

**Univerzita Karlova**  
**Přírodovědecká fakulta**

Studijní program: Speciální chemicko-biologické obory  
Studijní obor: Molekulární biologie a biochemie organismů



**MUDr. Lenka Daumová**

**Kodónové složení a exprese isoakceptorových tRNA jako mechanismus regulace genové exprese**

**Codon usage and isoacceptor tRNA expression as a mechanism to control gene expression**

Bakalářská práce

Školitel:

RNDr. Michal Čáp, PhD.

Praha, 2020



**Prohlášení:**

Prohlašuji, že jsem závěrečnou práci zpracovala samostatně a že jsem uvedla všechny použité informační zdroje a literaturu. Tato práce ani její podstatná část nebyla použita k získání jiného nebo stejného akademického titulu.

V Praze, dne 30.5.2020

podpis:

## **Poděkování**

Děkuji RNDr. Michalovi Čápovi, PhD. za cenné rady a vedení při psaní této bakalářské práce.

## Abstrakt

Genetický kód je definován jako klíč, podle kterého jsou k jednotlivým kodonům přiřazovány aminokyseliny. Genetický kód je degenerovaný, což znamená, že pro většinu aminokyselin existuje více synonymních kodonů. Po dlouhou dobu se myslelo, že takzvané tiché mutace, kdy nedojde k záměně aminokyseliny, ale pouze ke vzniku synonymního kodonu, nemají na genovou expresi vliv. Později se však podrobnějším výzkumem na molekulární úrovni zjistilo, že frekvence využití synonymních kodonů je jedním z faktorů ovlivňujících rychlost či efektivitu translace, stabilitu mRNA či schopnost proteinu zaujmout správnou konformaci. V celé řadě studií tak byl prokázán vliv použití určitých kodonů na genovou expresi.

Tato bakalářská práce je rešerší dostupné literatury týkající se výše zmíněných objevů. Nejprve jsou shrnuty základní informace ohledně tRNA, její struktury a modifikací v antikodonové smyčce. Dále je popsáno párování mezi kodonem a antikodonem a nekanonické párování wobble. Následně se soustředím na samotnou rozdílnou frekvenci používání synonymních kodonů, jejich preferencí (v anglicky psané literatuře označované termínem „codon bias“). Jsou rozebrány její možné příčiny, souvislost s obsahem GC párů v genomu, dohady o výzkumu či výběru optimálního kodonu a jaký je vliv používání synonymních kodonů na efektivitu translace. Jsou citovány mnohé studie na různých modelových organismech včetně člověka, které se snažily tento fenomén popsat.

Klíčová slova: exprese tRNA, efektivita translace, frekvence použití kodonů, modifikace tRNA, synonymní kodony

## Abstract

Genetic code is defined as a set of rules, which encode amino acid sequences in proteins, according to codon usage. It is widely known, that there are multiple codons for most amino acids, thanks to the degeneracy of the genetic code. There was a hypothesis, that silent mutations, which result in a synonymous codon and not in incorporating of a different amino acid into the peptide chain, don't affect the gene expression. Later however, it was found through more detailed research on molecular level, that codon usage bias is in fact one of the factors, that regulate translation effectivity and rate, mRNA stability or even protein folding and gene expression. There has been many studies published on these topics.

This bachelor thesis is a review of these studies. First, I summarize basic information on tRNA, its structure and modifications in anticodon loop. Next I write about base pairing between codon and anticodon, including the non-canonical wobble base pairing. Then I emphasize on codon bias, its causes, its relationship with genome GC content. I also include some author's conjectures about how to approach this phenomenon, which codons are optimal and what is the impact of codon usage bias on translation efficiency. I cite many studies on this topic, which was researched on many model organisms, including *Homo sapiens*.

Key words: tRNA expression, translation efficiency, codon usage bias, tRNA modifications, synonymous codons

## Seznam použitých zkratek

| <b>Zkratka</b>   | <b>Význam</b>                                 |
|------------------|---|
| A                | Adenin  |
| AMP              | Adenosin mono fosfát                          |
| ala              | Alanin  |
| araC             | Protein regulující transkripci araBAD operonu |
| arg              | Arginin                                       |
| asn              | Asparagin                                     |
| asp              | Kyselina asparagová                           |
| ATP              | Adenosin tri fosfát                           |
| C                | Cytosin                                       |
| CAI              | Codon adaptation index                        |
| Cm               | 2'-O-methyl-cytidin                           |
| cys              | Cystein                                       |
| DNA              | Deoxyribonukleová kyselina                    |
| elp3             | Elongator protein 3                           |
| fol              | Dihydrofolát reduktáza                        |
| FOP              | Frekvence optimálních kodonů                  |
| G                | Guanin  |
| GC               | Guanin Cytosin                                |
| gln              | Glutamin                                      |
| glu              | Kyselina glutamová                            |
| gly              | Glycin  |
| Gm               | 2'-O-methyl-guanosin                          |
| GTP              | Guanosin tri fosfát                           |
| his              | Histidin                                      |
| I                | Inosin  |
| i <sup>6</sup> A | N6-isopentenyl-adenosin                       |
| ile              | Izoleucin                                     |
| lacY             | permeáza laktosy                              |

|                                   |   |
|-----------------------------------|---|
| leu                               | Leucin  |
| lys                               | Lysin   |
| mcm <sup>5</sup> s <sup>2</sup> U | 5-methyl-karboxyl-methyl-2-thio uridin  |
| mcm <sup>5</sup> U                | 5-methyl-karboxyl-methyl uridin   |
| m <sup>1</sup> G                  | 1-methyl-guanosin   |
| m <sup>1</sup> I                  | 1-methyl-inosin   |
| m <sup>3</sup> C                  | 3-methyl cytidin  |
| m <sup>5</sup> C                  | 5-methyl cytidin  |
| ncm <sup>5</sup> U                | 5-karbamoyl-methyl-uridin   |
| ncm <sup>5</sup> Um               | 5-karbamoyl-methyl-2'-O-methyl uridin   |
| nTE                               | Normalized translational efficiency   |
| ompA                              | Outer membrane protein II   |
| phe                               | Fenylalanin   |
| pro                               | Prolin  |
| RNA                               | Ribonukleová kyselina   |
| RSCU                              | Relative Synonymous Codon Usage   |
| ser                               | Serin   |
| tAI                               | tRNA adaptation index   |
| t <sup>6</sup> A                  | N6-threonyl-karbamoyl-adenosin  |
| thr                               | Threonin  |
| thrA                              | Bifunkční enzym Aspartokináza I / homoserin dehydrogenáza I                       |
| trpC                              | Bifunkční enzym N-5'-fosforibosyl-anthranylát izomeráza / indol-3-glycerol fosfát |
| tufA                              | Elongační faktor Tu   |
| tyr                               | Tyrosin   |
| U                                 | Uracil  |
| val                               | Valin   |
| yW                                | Wybutosin   |
| ψ                                 | Pseudouridin  |



## Obsah

|       |  |    |
|-------|--|----|
| 1     | Úvod.....  | 1  |
| 2     | Transferová RNA .....  | 3  |
| 2.1   | Geny kódující tRNA .....   | 3  |
| 2.2   | Struktura tRNA.....  | 4  |
| 2.2.1 | Modifikace antikodonové smyčky .....   | 5  |
| 3     | Interakce kodonu a antikodonu.....   | 7  |
| 3.1   | Wobble párování.....   | 8  |
| 4     | Preference kodonů (codon bias).....  | 9  |
| 4.1   | Metody stanovení míry preference kodonů .....                                  | 10 |
| 4.2   | Stanovení optimálního kodonu.....  | 12 |
| 4.3   | Důvody pro preferenci synonymních kodonů .....                                 | 13 |
| 4.3.1 | Preferovaná báze na třetí pozici v kodonu a GC obsah v genomu .....            | 13 |
| 4.4   | Frekvence používání synonymních kodonů při expresi heterologních proteinů..... | 14 |
| 5     | Preference STOP kodonů .....   | 14 |
| 6     | Vliv používání synonymních kodonů na efektivitu translace.....                 | 16 |
| 6.1   | Výzkum preference kodonů a efektivitu translace ve všech doménách života ..... | 17 |
| 6.2   | Používání synonymních kodonů a efektivita translace .....                      | 18 |
| 6.2.1 | Efektivita translace - srovnání vybraných eukaryotických genomů.....           | 19 |
| 6.3   | Preference kodonů v závislosti na počtech kopií mRNA.....                      | 20 |
| 6.4   | Vliv preference kodonů na přesnost translace .....                             | 21 |
| 7     | Vliv použití synonymních kodonů na strukturu proteinu .....                    | 23 |
| 7.1   | Vliv používání vzácných kodonů na strukturu proteinu.....                      | 24 |
| 8     | Změny genové exprese ovlivněné preferencí kodonů či nadbytkem tRNA.....        | 24 |
| 9     | Preference kodonů u člověka .....  | 25 |
| 10    | Závěr.....   | 27 |
| 11    | Seznam použité literatury .....  | 28 |

## 1 Úvod

Molekula transferové RNA (tRNA) je jeden z typů ribonukleové kyseliny (RNA) vyskytující se v buňkách. Její funkce se uplatňuje při translaci, neboť na základě správné interakce kodonu a antikodonu přináší do nově se tvořícího polypeptidového řetězce správnou aminokyselinu.

Translace je proces, při kterém podle genetické informace vzniká molekula proteinu. Genetická informace je myšlena jako informace, kódovaná určitým genem v DNA, jejíž přepis do RNA dle komplementarity bází se nazývá transkripce. Na začátku transkripce je díky transkripčním faktorům nalezen promotor (TATA či CAT box) a enzymem DNA-dependentní RNA polymeráza je vytvořena molekula pre-mRNA, která je u eukaryot dále upravována, např. sestřihem intronů. Vzniká molekula messengerové RNA (mRNA).

Ta je v procesu translace „čtena“ mezi velkou a malou podjednotkou ribosomu (40S a 60S u eukaryot, 30S a 50S u prokaryot) po tripletech bází, které se nazývají kodony. Dekódování na základě komplementarity bází mezi kodonem a antikodonem (trojice bází, párující s kodonem) provádí molekula transferové RNA, která má na 2' či 3' –OH skupinu ribózy adenosinu v sekvenci 5'-CCA-3' na svém 3' konci navázanou aminokyselinu. Tuto reakci (aminoacylaci) katalyzují aminoacyl-tRNA-syntetázy, kterých je 20 druhů, pro každou aminokyselinu vlastní.

V ribosomu jsou tak zvaná A- P- a E- místa. Aminoacylovaná tRNA vstupuje nejprve do A-místa (aminokyselinového), kde probíhá čtení kodonu. V případě zařazení správné tRNA dojde k posunutí podjednotek ribosomu po mRNA o tři báze směrem ke 3' konci, posunutí právě čteného kodonu s navázanou tRNA s aminokyselinou do P-místa (peptidového). Do A-místa vstupuje další tRNA s aminokyselinou, a v případě komplementarity bází mezi kodonem a antikodonem zde opět zůstává a dochází k posunutí ribosomu výše popsaným způsobem. Pokud mezi bázemi není možnost párování, aminoacylovaná molekula tRNA vystupuje a přichází jiná tRNA s jiným antikodonem.

V ribosomu pak mezi A- a P- místem dochází mezi molekulami aminokyselin ke vzniku peptidové vazby, konkrétně mezi uhlíkem C1 na na C-konci polypeptidu a dusíkem N2 přistupující aminokyseliny. Ribosom se opět posune směrem ke 3' konci mRNA, tRNA z P-místa se dostává do E-místa (exit). Na nový kodon v A-místě nasedá další tRNA.

Proces translace pokračuje, dokud jsou na mRNA kodony, kódující některou z aminokyselin. Ukončuje ji STOP kodon (UAA, UAG či UGA), který způsobí rozložení celého komplexu ribosomu a mRNA pomocí

navázání release faktorů. Vzniklý protein se skládá a případně prochází dalšími post-translačními úpravami, v závislosti na tom, o jaký protein se jedná, jaká je jeho funkce či cílová lokalizace.

Pro většinu aminokyselin však existuje více kodonů, které ji kódují. Genetický kód je proto popisován jako degenerovaný. Molekuly tRNA, které váží stejnou aminokyselinu, ale obsahují rozdílný antikodon, se nazývají isoakceptorové. Navíc může jedna molekula tRNA dekódovat více kodonů díky takzvanému wobble párování. Jde o mechanismus, kdy (například kvůli modifikacím báze na wobble pozici v antikodonu) může docházet ke tvorbě vodíkových můstků i mezi jinými funkčními skupinami a tedy jinými bázemi, než by odpovídalo kanonickému párování mezi bázemi (Watson-Crickovo párování). Modifikace nukleotidů v tRNA tedy zvyšují možnosti párování.

Přestože jsou všechna tato fakta dlouhou dobu známá, teprve výzkumy z posledních let ukazují, že buňky a organismy mohou využívat různé kodonové složení jednotlivých genů např. k regulaci genové exprese. Vlivem mnoha faktorů popsaných dále mohou využívat různé kombinace synonymních kodonů v té samé mRNA, aby se zvýšila translační efektivita či růstová rychlost. Výzkum proběhl na genomech všech tří domén života (prokaryota, eukaryota i archea). U mnohobuněčných organismů bylo popsáno též rozdílné používání kodonů v rámci jednotlivých tkání.

Tako bakalářská práce má za úkol shrnout současné vědomosti týkající se frekvence používání synonymních kodonů a jejich preferenci, a jak tato ovlivňuje genovou expresi.

## 2 Transferová RNA

Transferová RNA se skládá ze 75 nukleotidového jádra a 3'-CCA konce (Pak *et al.*, 2017), počet nukleotidů se však může pohybovat mezi 73 a 95 (Anderson & Ivanov, 2014). Přibližná molekulová hmotnost tRNA je 25 000 Da (Clark, 2006).

Transferová RNA se v buňce vyskytuje ve velkém nadbytku a bývá výrazně modifikovaná; průměrný počet modifikací na molekulu je 13 (Pan, 2018). Primární transkript obsahuje sice pouze standardní báze A, C, G, U, avšak v průběhu maturace dochází k jejich četným modifikacím. Z celkového počtu modifikací různých typů RNA bývá nejvíce modifikována právě tRNA. Četnost se liší i napříč různými organismy nebo je rozdílná u jaderné a například plastidové tRNA. Modifikace ovlivňují molekulární vlastnosti, jako je například výsledná terciární struktura či adaptace k různým teplotám prostředí (u termofilních organismů brání denaturaci molekuly zvýšením její stability, u psychrofilních organismů zase zvyšují flexibilitu struktury) (shrnuto v: Lorenz *et al.* (2017)). Přehled těchto změn uvádí databáze „The RNA Modification Database“ (Cantara *et al.*, 2011).

### 2.1 Geny kódující tRNA

Molekula tRNA vzniká transkripcí svých genů za pomoci komplexu RNA polymerázy III. Množství jednotlivých isoakceptorů v buňce je tedy regulováno jejich transkripcí a počtem jejich genů (shrnuto v: Novoa & Ribas de Pouplana (2012)). Počet genů pro každou tRNA se liší a souvisí s používanými kodony. Například na kvasince *Saccharomyces cerevisiae*, konkrétně kmeni S288C, bylo nalezeno 274 genů pro tRNA. Z těchto následně vzniká 42 druhů tRNA s různými antikodony (Percudani *et al.*, 1997).

U mnohobuněčných organismů (například člověka) se exprese různých isoakceptorových tRNA liší v závislosti na typu tkáně (Dittmar *et al.*, 2006). Autoři porovnávali expresi isoakceptorových tRNA v osmi různých tkáních. Pro detekci použili microarray metodu. Zjistili, že exprese jednotlivých typů tRNA se lišila tkáň od tkáně (dle používaných kodonů – viz také Plotkin *et al.* (2004)) a v mozku byla celkově vyšší (Dittmar *et al.*, 2006). Ve srovnání s mozkovou tkání bylo nalezeno například v orgánech imunitního systému větší množství molekul tRNA přenášejících hydrofobní aminokyseliny, těchto bylo zase v reprodukčních tkáních nalezeno méně. V mozku bylo nejvíce molekul tRNA obsahujících antikodonovou sekvenci pro většinu nabitých a polárních aminokyselin. Souvisí to pravděpodobně s dostupností aminokyselin v těchto tkáních.

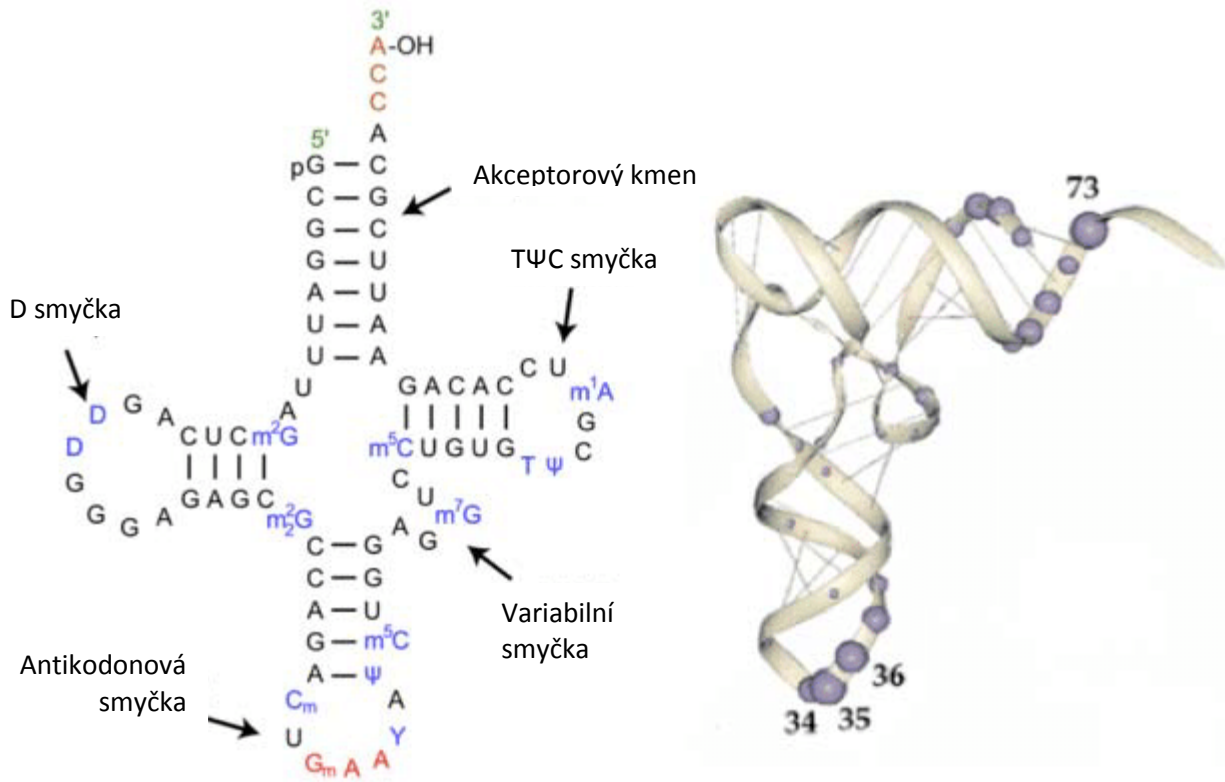
## 2.2 Struktura tRNA

Z hlediska sekundární struktury je možno popsat molekulu transferové RNA jako „jetelový list“ (v anglické literatuře označovaný jako „cloverleaf“), neboť dochází k typickému párování bází v rámci molekuly. „Jetelový list“ se skládá z několika částí (Lorenz *et al.*, 2017). Terciární strukturu tRNA lze popsat jako tvar písmene L. Obě struktury transferové RNA ukazuje **obrázek č.1**.

V molekule tRNA, jako obecně u nukleových kyselin, je popisován 5'- a 3'-konec. Tyto dohromady tvoří akceptorový stonek („acceptor stem“). Na 3'-konci se nachází triplet bází CCA. Tam je pomocí enzymů aminoacyl-tRNA-syntetáz připojena aminokyselina, jak je popsáno dále. Dále se dají rozlišit tzv. „loops“, smyčky, či „arms“, raménka – D-raménko, antikodonové raménko, variabilní smyčka a T-Ψ-raménko (shrnutí v: Lorenz *et al.* (2017)). Názvy ramének jsou odvozeny od zkratk modifikovaných bází, které se v nich nachází. Pro D-raménko je to dihydrouridin, v T-Ψ-raménku jsou ribothymidin a pseudouridin (Nobles *et al.*, 2002). Některé tyto modifikované nukleotidy se označují jako konzervované, neboť se na těchto konkrétních pozicích u molekul tRNA objevují s velmi vysokou frekvencí (Sprinzl *et al.*, 1998).

Ne všechny tRNA však zaujímají sekundární strukturu „jetelového listu“. Toto bylo popsáno například u mitochondriálních tRNA, které jsou kratší a můžou jim chybět některé části, například celá D- či T-raménka, nebo i obě. Tato raménka jsou důležitá pro tvorbu terciární struktury (shrnutí v: Lorenz *et al.* (2017)).

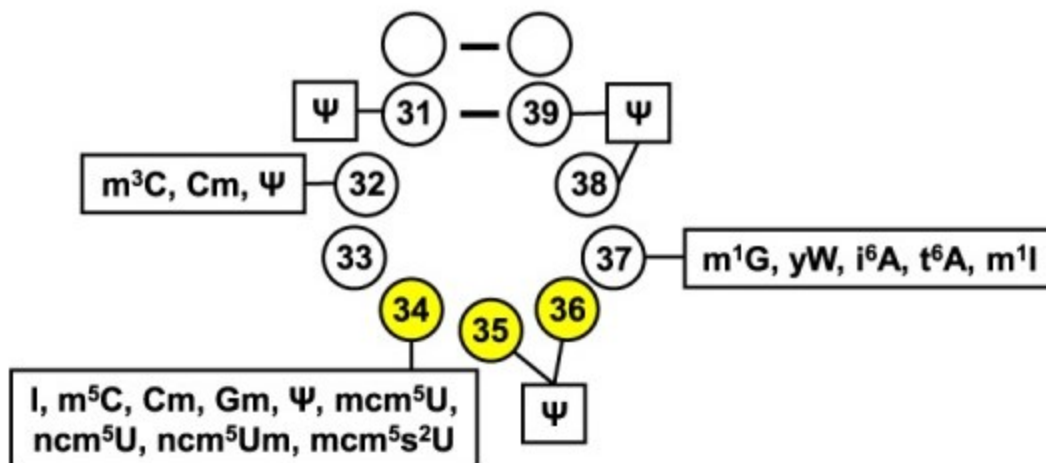
V rámci terciární struktury je na jednom konci antikodon, na druhém 3'-CCA konec. V ohybu struktury dochází k interakci D- a T-ramének (Pak *et al.*, 2017). Při těchto interakcích a skládání molekuly tRNA do terciární struktury se účastní přítomné modifikované nukleotidy a  $Mg^{2+}$  ionty (Nobles *et al.*, 2002).



**Obrázek č.1:** Struktura tRNA, vlevo sekundární, vpravo terciární. Vyznačeny jsou i některé modifikace bází, jejichž zkratky jsou uvedeny v seznamu použitých zkratk. V terciární struktuře je vyznačeno číslování bází antikodonu (34-36). Společně s bází na 73.pozici jsou tyto nukleotidy rozeznávány aminoacyl-tRNA-syntetázami. Upraveno podle: Hartman & Smith (2014).

### 2.2.1 Modifikace antikodonové smyčky

Modifikací se rozumí biochemická úprava báze pomocí různých enzymů – například přidání funkční skupiny. Příklady možných modifikací antikodonové smyčky uvádí **obrázek č.2**.



**Obrázek č.2:** Možnosti modifikací bází tRNA v antikodonové smyčce u kvasinky *Saccharomyces cerevisiae*. Čísla označují pořadí báze v molekule tRNA, žlutě vyznačené nukleotidy 34 až 36 tvoří antikodon. Význam zkratk je uveden výše v části Seznam použitých zkratk. Převzato z: Han & Phizicky (2018)

Modifikace ve wobble pozici mohou ovlivňovat, se kterou bází v kodonu bude možné párování. U různých organismů se mohou modifikace lišit (Higgs & Ran, 2008).

U antikodonové smyčky jsou modifikace dobře prozkoumané. Napříč organismy je zde největší variabilita. Je popsán jejich vznik v molekule v konkrétním pořadí jednotlivými modifikačními enzymy. U antikodonové smyčky se jednotlivé modifikace liší podle toho, o kterou tRNA se jedná, neboť jednotlivé druhy tRNA mají rozdílnou sekvenci v antikodonu. Některé báze mohou zůstat beze změn, z jiných mohou vzniknout až čtyři různé odvozené báze v závislosti na původní bázi (Han & Phizicky, 2018). Uridin ve wobble pozici (34) je vždy modifikovaný (Björk *et al.*, 2007).

Modifikace antikodonové smyčky v různých pozicích přímo ovlivňují na různých úrovních proces translace. Například bylo zjištěno, že pomáhají udržet správný čtecí rámec (Urbonavicius *et al.*, 2001). Ovlivňují i schopnost tRNA rozpoznávat synonymní kodony – například studie zabývající se krystalovými strukturami tRNA pro valin u bakterie *Escherichia coli* prokázala, že modifikací uridinu na pozici 34 se dá dosáhnout možnosti párování antikodonu se všemi čtyřmi možnými kodony pro valin (GUA, GUG, GUC, GUU) (Weixlbaumer *et al.*, 2007).

U kvasinky *Saccharomyces cerevisiae* bylo popsáno, že modifikace uridinu na wobble pozici na mcm<sup>5</sup>s<sup>2</sup>U v tRNA pro glutamin, kyselinu glutamovou a lysin je pro kvasinky nezbytná pro správný průběh

proteosyntézy. Björk *et al.* (2007) zkoumali vliv delece genů pro proteiny Tuc1p a Elp3p, které modifikují nukleotid U34. Tato modifikace je pro kvasinky esenciální, neboť dvojitá delece *TUC1* a *ELP3* je letální kvůli špatnému čtení mRNA a tedy poruše proteosyntézy (Björk *et al.*, 2007).

### 3 Interakce kodonu a antikodonu

Genetický kód obsahuje 61 kodonů, které kódují 20 aminokyselin a 3 kodony kódující STOP signál pro translaci, tzv. STOP kodony. Osmnáct aminokyselin je kódováno více kodony, což činí genetický kód degenerovaný. Jako univerzální ho nazýváme proto, že se napříč organismy téměř nemění (Quax *et al.*, 2015). Přehled, které aminokyseliny jsou kódovány jakými kodony (a kolik kodonů to pro každou aminokyselinu je), ukazuje **tabulka č.1**.

Z tabulky je patrné, že některé aminokyseliny jsou kódovány pouze jedním kodonem (methionin, tryptofan), ostatní dvěma až šesti různými kodony. U těchto „vícekodonových“ aminokyselin zpravidla nejsou všechny synonymní kodony využívány se stejnou frekvencí. Tento jev se nazývá preference kodonů (codon bias). Jeho možnými důvody, stejně jako jeho využitím pro regulaci genové exprese se zabýváme v dalších částech práce.

**Tabulka č.1:** Tabulka ukazující kódování jednotlivých aminokyselin pomocí kodonů. Kodony UAA, UAG a UGA znamenají ukončení translace. Upraveno podle: Sánchez *et al.* (2006). Aminokyseliny jsou zde popsány jednopísmennými symboly. P = prolin, L = leucin, I = isoleucin, M = methionin, V = valin, S = serin, T = threonin, A = alanin, Y = tyrosin, H = histidin, Q = glutamin, N = asparagin, K = lysin, D = asparagová kyselina, E = glutamová kyselina, C = cystein, W = tryptofan, R = arginin, G = glycin.



|                   |   | Druhá báze kodonu |   |     |   |      |     |      |   |   |
|-------------------|---|-------------------|---|-----|---|------|-----|------|---|---|
|                   |   | U                 |   | C   |   | A    |     | G    |   |   |
| První báze kodonu | U | UUU               | P | UCU | S | UAU  | Y   | UGU  | C | U |
|                   |   | UUC               |   | UCC |   | UAC  |     | UGC  |   | C |
|                   |   | UUA               | L | UCA |   | Stop | UGA | Stop | A |   |
|                   |   | UUG               |   | UCG |   |      | UAG | UGG  | W | G |
|                   | C | CUU               | L | CCU | P | CAU  | H   | CGU  | R | U |
|                   |   | CUC               |   | CCC |   | CAC  |     | CGC  |   | C |
|                   |   | CUA               |   | CCA |   | CAA  | CGA | A    |   |   |
|                   |   | CUG               |   | CCG |   | CAG  | CGG | G    |   |   |
|                   | A | AUU               | I | ACU | T | AAU  | N   | AGU  | S | U |
|                   |   | AUC               |   | ACC |   | AAC  |     | AGC  |   | C |
|                   |   | AUA               |   | ACA |   | AAA  | AGA | A    |   |   |
|                   |   | AUG               | M | ACG |   | AAG  | K   | AGG  | R | G |
|                   | G | GUU               | V | GCU | A | GAU  | D   | GGU  | G | U |
|                   |   | GUC               |   | GCC |   | GAC  |     | GGC  |   | C |
|                   |   | GUA               |   | GCA |   | GAA  | GGA | A    |   |   |
|                   |   | GUG               |   | GCG |   | GAG  | GGG | G    |   |   |

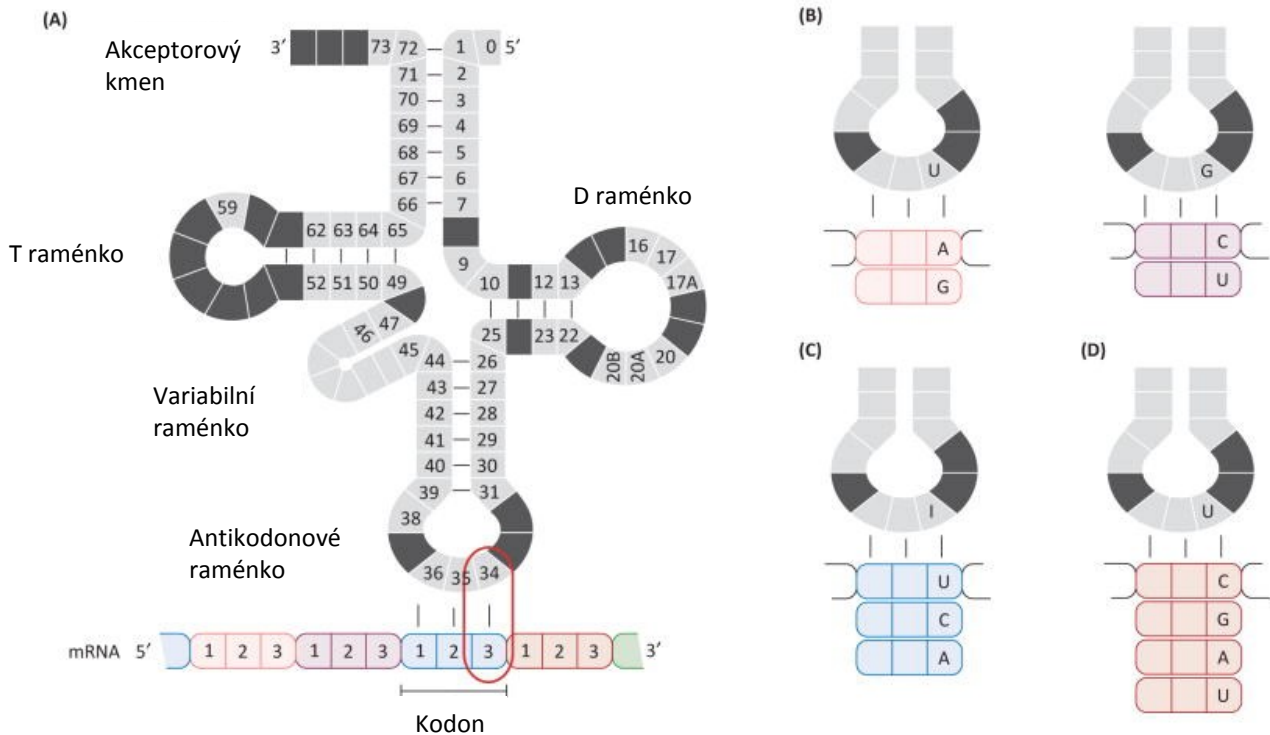
1 Jednopísmenná zkratka aminokyseliny

### 3.1 Wobble párování

Více synonymních kodonů (kódujících stejnou aminokyselinu) ale může být translatováno stejnou tRNA díky fenoménu zvanému wobble (Crick, 1966). Jedná se o jev, kdy při párování bází na třetí pozici kodonu jsou možné interakce mezi bázemi, které se vymykají Watson-Crickovu standardnímu párování. Crick identifikoval sedm různých možností, které mohou nastat při wobble párování a popsal i možnosti párování, kterých není možné dosáhnout (Crick, 1966).

K wobble dochází na pozici 34 v tRNA, což je báze v první pozici na antikodonu (v kodonu na mRNA páruje s bází na třetí pozici) – zobrazeno na **obrázku č.3**. Například uridin v těchto místech u eukaryot bývá modifikovaný připojením funkční skupiny či různých postranních řetězců. Je tak možné, aby jedna tRNA dekodovala více kodonových tripletů (Johansson *et al.*, 2008). Pokud je báze na wobble pozici schopná

párovat s jakoukoliv bází v první pozici kodonu, je tento jev nazýván superwobble, což bylo popsáno například v případě uridinu v pozici 34 u chloroplastových tRNA (Alkatib *et al.*, 2012).



**Obrázek č.3:** Wobble párování mezi nukleotidy 3.pozice kodonu a 1.pozice antikodonu (nukleotid č.34 v molekule tRNA). **(A)** V pravé části obrázku jsou vyznačeny možnosti párování bazí podle wobble či Watson-Crickova párování **(B) (C)** či superwobble **(D)**. A = adenin, C = cytosin, I = inosin, G = guanin, U = uracil. Tmavě vyznačené pozice tRNA jsou konzervované. Upraveno podle: Mauro & Chappell (2014)

Pomocí wobble lze tedy dosáhnout toho, že kodon bude dekódován jinou isoakceptorovou tRNA. Avšak v případě, že je k dispozici tRNA, jejíž antikodon bezchybně páruje s kodonem, je tato využívána pro translaci přednostně a wobble párování isoakceptorové tRNA se neuplatňuje (Percudani *et al.*, 1997).

#### 4 Preference kodonů (codon bias)

Termínem preference kodonů je označována odlišná frekvence používání jednotlivých synonymních kodonů pro kódování té samé aminokyseliny. Jak bylo popsáno v mnohých studiích, použití kodonů, které

jsou využívány tou isoakceptorovou tRNA, jíž je v buňce větší množství, je spojeno s efektivnější translací a tedy s vyšší expresí proteinu (Berg & Kurland, 1997) (Hershberg & Petrov, 2008) (Tuller *et al.*, 2010).

U příbuzných druhů organismů jsou běžně preferovány podobné synonymní kodony, liší se však frekvence použití kodonů mezi různými druhy (Vicario *et al.*, 2007). Nicméně genetický kód pro jednotlivé aminokyseliny se mezi organismy nemění (Fredrick & Ibba, 2010).

O preferenci používání některého ze synonymních kodonů se mluví již poměrně dlouho. Jedna z časných studií publikovaných na toto téma byla provedena na modelovém organismu *Escherichia coli* pro sedm různých genů (ompA, tufA, fol, lacY, thrA, trpC a araC) (Ikemura, 1981). Korelace mezi množstvím konkrétní tRNA a jejím využitím byla analyzována pomocí lineární regrese. Autoři se pokusili stanovit, které kodony jsou pro *E. coli* optimální, společně s nadbytkem příslušné tRNA. Účinnost translace se totiž zvyšuje při užívání právě těch kodonů, pro které je příslušná molekula tRNA v nadbytku (Ikemura, 1981).

S častějším využíváním upřednostňovaných kodonů se setkáváme i u bakteriálních virů, bakteriofágů, které neexprimují vlastní tRNA. V průběhu evoluce se adaptovaly na používání translačního aparátu buňky, kterou infikují. Míra této adaptace je možná stanovit pomocí indexu adaptace kodonů (CAI, viz dále) (Sharp & Li, 1987).

V **příloze č.1** jsou tabulky, ukazující četnosti využívání jednotlivých kodonů a zastoupení GC párů v nich u bakterie *Escherichia coli*, kvasinky *Saccharomyces cerevisiae* a člověka (*Homo sapiens*). Z tabulek je patrné, že tyto organismy preferují různé kodony a mají rozdílný obsah GC párů v genomu i na jednotlivých pozicích v kodonech. Toto je diskutováno dále.

#### 4.1 Metody stanovení míry preference kodonů

Pro výzkum frekvence používání synonymních kodonů jsou užívána data získaná sekvenací genomu a analyzována pomocí statistických metod a výpočtů; jsou tvořeny matematické vzorce umožňující analýzy kodonové preference *in silico*. Je možné stanovit míru zastoupení jednotlivých synonymních kodonů v určitém genu a porovnávat frekvenci kodonů v celém genomu daného organismu; či je porovnávat dle míry zastoupení bází s celým genomem (například dle obsahu GC párů), s počtem genů pro příslušné isoakceptorové tRNA (jejich nadbytkem), či srovnávat preferenci kodonů v genomech různých organismů.

Dále uvádím některé metody stanovení preference kodonů, se kterými jsem se setkala v publikacích při psaní této bakalářské práce.

Metoda zvaná RSCU (Relative Synonymous Codon Usage) (Sharp *et al.*, 1986) udává podíl pozorované četnosti synonymního kodonu a očekávané četnosti, pokud by k preferenci mezi kodony nedocházelo. Tento výpočet je možné použít také k porovnávání kodonové preference mezi různými organismy, či mezi různými geny.

Pro měření míry používání synonymních kodonů se zohledněním evolučních mechanismů (přirozená selekce, mutace) byl již v 80. letech vyvinut tzv. Codon Adaptation index (CAI). Jeho hodnoty souvisí s mírou exprese konkrétního proteinu. Málo exprimované proteiny měly nízké hodnoty CAI (Sharp & Li, 1987). Na této metodě byl následně založen software Cod4Bio, s jehož pomocí je možné odhadnout expresi heterologních rekombinantních proteinů (Mignon *et al.*, 2018).

Byl vytvořen matematický model, popisující vztah množství určité tRNA a frekvenci využívání synonymních kodonů z evolučního hlediska (autor vzal do úvahy mutační rychlost a velikost populace u rychle se množících jednobuněčných organismů – *E. coli* a kvasinek). V jejich genech se častěji vyskytují kodony, které jsou rozeznávány molekulami tRNA, kterých je v buňce větší množství (Bulmer, 1987).

Existuje databáze frekvence použití jednotlivých kodonů pro různé organismy, která je dostupná na internetu (<http://www.kazusa.or.jp/codon/>) (Nakamura, 2000).

Adaptační index tRNA (tAI) zohledňuje počet kopií genů pro tRNA schopných přeložit daný kodon. Do úvahy je přitom bráno i wobble párování bází (dos Reis, Savva, & Wernisch, 2004). Index tAI původně vypracovaný pro kvasinkové tRNA byl následně stanoven i pro další organismy (Sabi & Tuller, 2014).

Výpočet frekvence optimálních kodonů (FOP) udává, s jakou četností se v rámci synonymních kodonů pro tutéž aminokyselinu vyskytuje kodon s největším počtem genů pro příslušnou tRNA (Lavner & Kotlar, 2005).

Index označovaný jako %MinMax udává frekvenci používání jednotlivých kodonů a určuje nakolik je daná sekvence mRNA (tedy kodonů) běžná či vzácná (Clarke IV & Clark, 2008).

Byla popsána statistická metoda, zkoumající tzv. normalizovanou translační efektivitu (normalized translation efficiency, nTE) pro lepší srovnání různých organismů (Pechmann & Frydman, 2013). Tato statistika říká o kodonech, že jsou optimální, pokud je v genomu dostatečný nadbytek genů pro tRNA s příslušným antikodonem. Aminokyseliny jsou však z tohoto hlediska kódovány jak optimálními, tak méně optimálními kodony.

## 4.2 Stanovení optimálního kodonu

Jak již bylo zmíněno výše, byla vyvinuta řada matematických modelů pro určení, které kodony jsou u kterého organismu z hlediska translační efektivity optimální. Ne vždy se však výsledky zjištěné různými studii shodovaly, což dokazuje například následující debata dvou vědeckých skupin.

Studie provedená Hershberg a Petrovem v roce 2009 se zabývala určením pravidel pro výběr optimálního kodonu na souboru 675 bakteriálních genomů, 52 archeálních a 10 genomů hub. Podle obsahu jednotlivých nukleotidů byli schopni u každého genomu popsat, které kodony jsou preferovány. U genomů, kde byl vyšší obsah GC, byly jako optimální identifikovány kodony s vyšším obsahem GC a naopak. Dále identifikovali optimální kodon pro každou aminokyselinu dle počtu jejích kodonů a GC obsahu genomu (Hershberg & Petrov, 2009).

Wang *et al.* (2011) s těmito závěry nesouhlasili, neboť dle jejich názoru Hershberg & Petrov (2009) ve své studii nezohlednili selekci kodonů v průběhu evoluce a použili pouze korelační metodu bez porovnání spolehlivosti predikovaných kodonů. Wang *et al.* (2011) vyzkoušeli obě metody (korelaci a porovnání) na 203 bakteriálních kmenech, z nichž polovina měla vysoký počet kopií genů pro tRNA (s vysokým vlivem selekce na používané kodony, například geny pro ribosomální proteiny) a druhá nízký. Výsledky této studie ohledně toho, které kodony jsou optimální (i z hlediska obsahu GC párů v genomu) se od předchozí studie provedené skupinou Hershberg & Petrov (2009) značně lišily (Wang *et al.*, 2011).

Avšak Hershberg a Petrov v následujícím výzkumu porovnali obě metody a potvrdili platnost svých předchozích výsledků. Zároveň diskutovali vhodnost používání ribosomálních proteinů pro výzkum různé frekvence používání synonymních kodonů, neboť tyto geny jsou při výzkumu často používány. Avšak i ostatní geny podléhají selekci pro optimalizaci translace a používání kodonů a mělo by to být při výzkumu zohledněno (například při vzájemném srovnávání použitých kodonů v různých genech) (Hershberg & Petrov, 2012).

Některé studie vysvětlovaly použití ribosomálních genů (jako příkladu vysoce exprimovaných genů) tak, že jiné geny jsou v různých tkáních organismů exprimovány v různé míře. Někdy se však proto výsledky nemusí shodovat s dalšími studii (Hiraoka *et al.*, 2009).

### 4.3 Důvody pro preferenci synonymních kodonů

Proč se objevuje fenomén různé frekvence používání synonymních kodonů, se snažila objasnit řada studií. Často popisovanými příčinami byly selekční mechanismy (Sharp & Li, 1987) a bodové mutace (Kimura, 1980). Existuje souvislost mezi evolucí tRNA genů a četností používaných kodonů. Bakterie, které se rychle množí, potřebují vysokou rychlost translace. Díky evolučním mechanismům (například selekci) se u nich v genomu vyskytuje i více genů pro příslušné tRNA (Higgs & Ran, 2008).

Další důvod pro rozdílné používání synonymních kodonů může být například obsah GC párů v genomu, který se může u různých organismů výrazně lišit (Chen *et al.*, 2004). Dle vytvořeného modelu existuje souvislost mezi používáním kodonů a genomovým obsahem GC, takže organismy s vyšším podílem GC párů budou používat více kodony s vyšším obsahem GC (Knight *et al.*, 2001).

V genomu je možné identifikovat takzvané izochory, což jsou úseky DNA lišící se vzájemně obsahem GC párů (Bernardi *et al.*, 1985). Jeden z mechanismů, jak se může obsah GC párů v genomu zvyšovat, je například genová konverze (kopie homologního úseku při rekombinaci), pokud tato není opravena, či je opravena ve prospěch GC párů (Galtier *et al.*, 2001). Mechanismů evoluce množství GC párů v genomu však může být i více, dosud neprobádaných.

Výběr kodonů souvisí také se sekundární strukturou mRNA. U *Saccharomyces cerevisiae* bylo popsáno, že v některých genech jsou upřednostňovány takové kodony, které umožní vznik sekundárních struktur (například dvouřetězcových úseků, vláseček a smyček) na mRNA. Tyto sekundární struktury mohou mít funkci v regulaci sestřihu a stability mRNA. Kodonové složení ovlivňuje také strukturu pre-mRNA (než dojde k vystřížení intronů) z hlediska zvýšení termodynamické stability molekuly (Trotta, 2013); také je dokumentován vliv kodonového složení na míru degradace výsledné mRNA (Presnyak *et al.*, 2015).

#### 4.3.1 Preferovaná báze na třetí pozici v kodonu a GC obsah v genomu

Obsah GC párů genomu má někdy vliv i na výběr báze na třetí pozici v optimálním kodonu u daného organismu, který bude častěji využíván.

Na genomu 1346 prokaryotních organismů byla provedena studie, která zkoumala frekvenci použití kodonů končících pyrimidinovou bází (U či C). Byly to kodony, pro které nebyl nadbytek příslušné tRNA, jejich výběr tedy nebyl ovlivněn tímto způsobem. Byla zjištěna souvislost mezi počtem kodonů dané aminokyseliny a preferovanou bází ve třetí pozici kodonu. Serin, lysin a arginin, aminokyseliny kódované

nejvíce kodony, měly častěji na konci kodonu U, zatímco u aminokyselin se dvěma možnými kodony byl spíše použit kodon s C na třetí pozici. Důvodem pro tento nepoměr by mohla být lepší interakce během párování kodonu s antikodonem (U ve třetí pozici kodonu páruje s G na první pozici antikodonu méně „lepivě“, než v případě C ve třetí pozici kodonu, což zlepšuje jejich interakci během translace) a tím pádem je menší šance chybného zařazení aminokyseliny v proteosyntéze. V některých případech mělo vliv i množství GC párů v genomu (Wald *et al.*, 2012).

#### 4.4 Frekvence používání synonymních kodonů při expresi heterologních proteinů

Jelikož různé organismy preferují různé synonymní kodony, optimalizování použitých kodonů se využívá v genomovém inženýrství, aby se zvýšila efektivita produkce heterologních proteinů (Gustafsson *et al.*, 2004). Toto se využívá například ve farmaceutickém průmyslu či pro produkci jiných látek jako enzymů a dalších rekombinantních proteinů (Elena *et al.*, 2014).

Existuje řada aplikací a programů, pomocí nichž je možné *in silico* modelovat, které kodony jsou daným mikroorganismem preferované, aby jejich optimalita co nejvíce odpovídala translačnímu aparátu konkrétního vybraného mikroorganismu, a tak se co nejvíce zvýšila výtěžnost produkce heterologních proteinů (Puigbò *et al.*, 2007) (Chin *et al.*, 2014) (Daniel *et al.*, 2015).

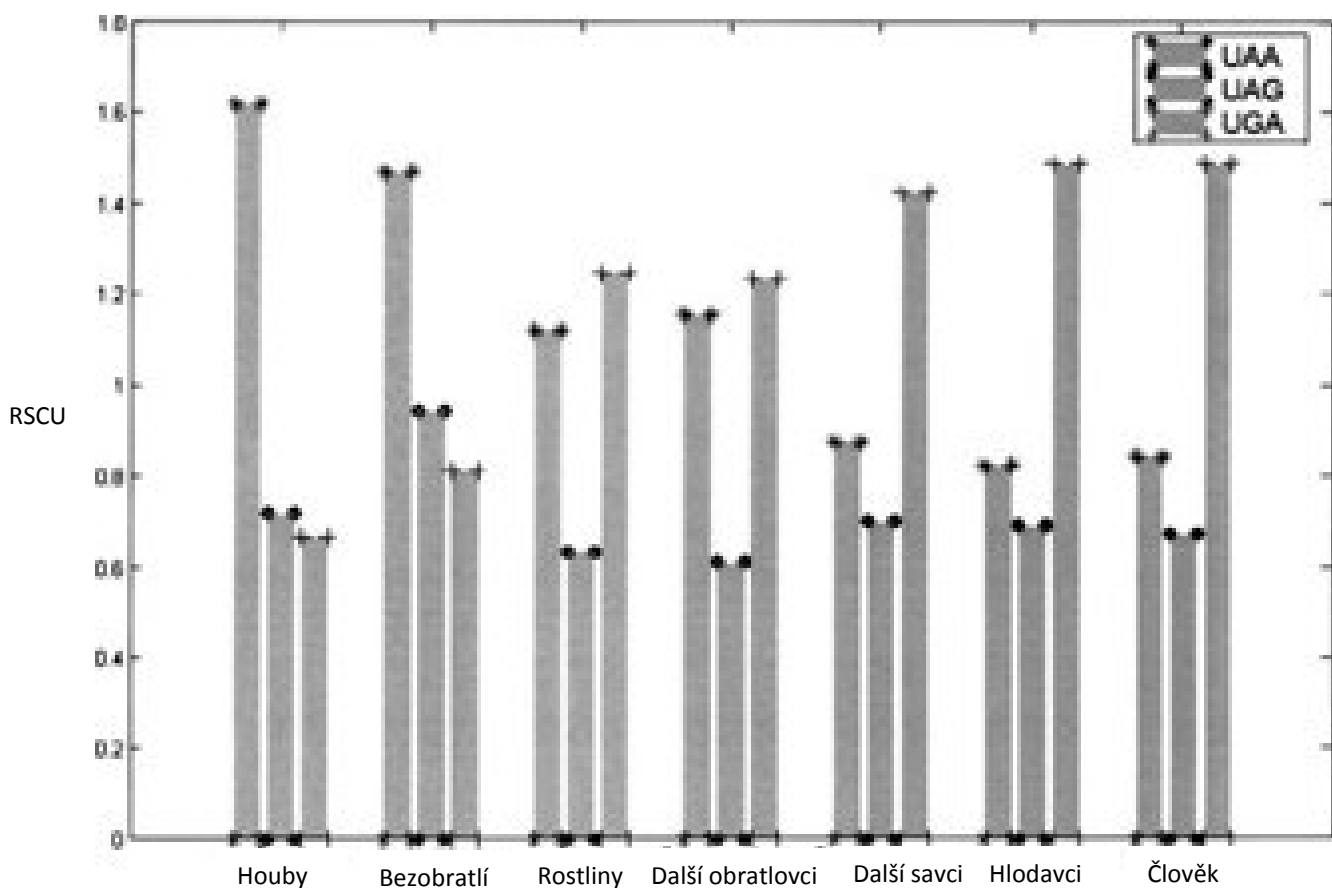
Byly provedeny i *in vivo* experimenty. Například skupina Panicker *et al.* (2015) se snažila zkoumat vliv použití alternativního START kodonu (místo standardního AUG například GUG) na genovou expresi heterologního proteinu u *Mycoplasma gallisepticum*. Inicie translačního procesu byla v případě GUG více efektivní, avšak na úkor jeho účinnosti (Panicker *et al.*, 2015).

## 5 Preference STOP kodonů

Pro terminaci proteosyntézy existují tři kodony (UAG, UAA a UGA). Jelikož je jejich efekt ve výsledku stejný (a jsou většinou tedy také považovány za synonymní), některé studie se zaměřily i na výzkum frekvence používání těchto STOP kodonů. Analýzou eukaryotických genomů bylo zjištěno, že i STOP kodony jsou využívány s různou frekvencí. U různých organismů se preferovaný STOP kodon liší, avšak nebyl potvrzen vztah k obsahu GC párů v genomu ani k úrovni genové exprese. Záleželo však na sekundární struktuře transkriptu, například UAA se spíše vyskytoval se smyčkách (Sun *et al.*, 2005).

Kodon UAA je ve větší míře používán nižšími eukaryoty (například kvasinky či bezobratlí), UGA u eukaryot s větší komplexitou (obratlovci). Poměr mezi zastoupením těchto dvou STOP kodonů se postupně mění s rostoucí komplexitou organismu. UAG se ukázal v porovnání s předchozími jako všeobecně nejméně používaný. Avšak u genomů rostlin a některých obratlovců nebyla zjištěna preference UAA před UGA (Sun *et al.*, 2005). Výsledky jsou zobrazeny na **obrázku č.4**.

U člověka je častější preference UGA vysvětlována vyšším obsahem GC párů v kódujících sekvencích a obecně v genomu, avšak v AT-bohatých oblastech je preferován UAA (Trotta, 2016).



**Obrázek č.4** Zastoupení STOP kodonů ve vybraných eukaryotických genomech. Vysvětlení je popsáno výše v textu. Na ose Y leží hodnoty RSCU (viz výše). Upraveno podle: Sun *et al.* (2005)

Někdy však kodon UGA může kódovat tryptofan, jak bylo popsáno u *Mycoplasma*. UGA i UGG kódují tryptofan také v mitochondriálním genomu (Yamao *et al.*, 1985).



Pomocí *in vivo* genetických modifikací lze dosáhnout toho, že STOP kodon bude alternativně smysluplně translatován, například některou z nestandardních aminokyselin. U kmene MG1655 bakterie *Escherichia coli* autoři změnili pomocí *in vivo* genomové editace funkci STOP kodonu UAG (který je v tomto genomu nejméně častý) delecí release faktoru 1, rozeznávajícího právě tento vzácný STOP-kodon. Za pomoci nově vložené aminoacyl-tRNA syntetázy mohla být jeho čtením do peptidového řetězce vnesena nestandardní aminokyselina (zde konkrétně p-azidofenylalanin, 2-naftalalanin či fosfoserin). Jako STOP kodon byl používán UAA, který využívá release faktor 2, jenž byl v buňce normálně přítomen. Ačkoli se objevila zvýšená odolnost vůči fágovi T7 (neboť 6 z jeho 60 STOP-kodonů jsou UAG a v buňce chyběl release faktor 1), někdy však takováto záměna snížila fitness bakteriální buňky, například v případě vnesení fosfoserinu docházelo ke špatnému skládání tvořených proteinů (Lajoie *et al.*, 2013).

Bylo také zjištěno, že existují preferované kodony, používané za start kodonem AUG a před jedním ze STOP kodonů, v závislosti na obsahu CG párů v genomu (Paul *et al.*, 2018).

## 6 Vliv používání synonymních kodonů na efektivitu translace

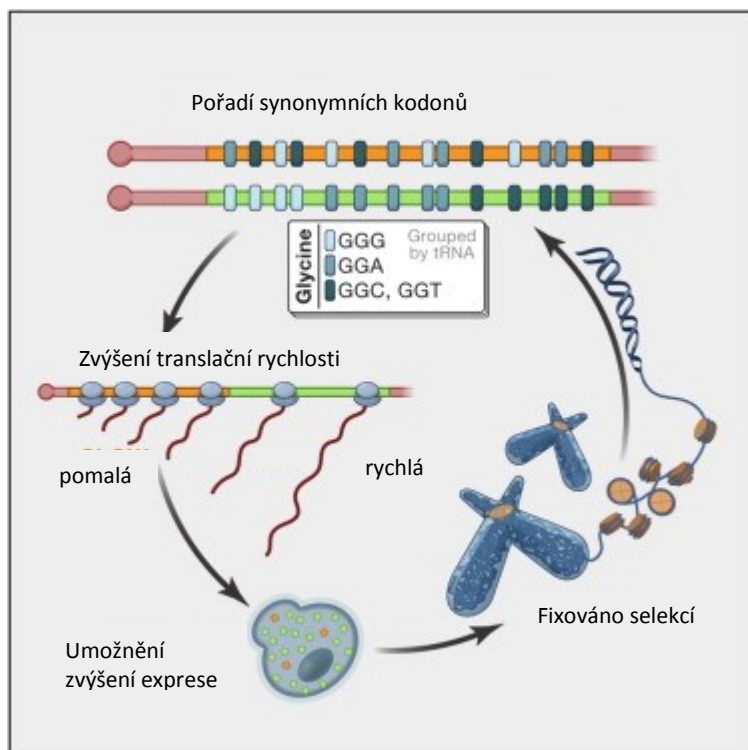
Již dlouho je známo, že (například díky evolučním procesům jako selekce, mutace či drift) preference mezi kodony a optimalizování translačního procesu z hlediska translační rychlosti či přesnosti u jednobuněčných organismů zvyšuje jejich růstovou rychlost a fitness (Bulmer, 1991). Rychlost translace se obecně liší u prokaryot a eukaryot (u bakterií je rychlejší) (Cannarrozzi *et al.*, 2010).

Z hlediska určení translační rychlosti je limitující krok její iniciace. Tato může být ovlivněna také použitím preferovaných synonymních kodonů na počátku transkriptu, neboť se rychleji translatují (Chu *et al.*, 2014). Používání optimálních kodonů, na rozdíl od těch méně preferovaných, také zvyšuje rychlost elongace translace (Yu *et al.*, 2015).

Která část transkriptu je právě čtena (a tedy který přesně kodon je právě používán), lze zjistit pomocí metody zvané „ribosome profiling“ (Ingolia *et al.*, 2009). S její pomocí je možné určit i jak dlouho je kodon přítomný v ribosomu na A- či P-místě, a tedy odhadnout jeho optimalitu na základě rychlosti jeho translace (Brule & Grayhack, 2017).

## 6.1 Výzkum preference kodonů a efektivity translace ve všech doménách života

Cannarrozzi *et al.* (2010) zkoumali genom kvasinky *Saccharomyces cerevisiae*, konkrétně translaci devíti „vícekodonových“ aminokyselin (ala, arg, gly, ile, pro, leu, ser, thr, val). Byla porovnávána frekvence použitých synonymních kodonů s následujícím kodonem pro tu samou aminokyselinu. Byla zjištěna vzájemná závislost, neboť tyto kodony jsou často translatovány tou samou molekulou tRNA, jak ukazuje **obrázek č.5**. Toto bylo pozorováno pro časté i vzácné kodony, a výrazněji u genů, které byly více exprimované. S větší vzdáleností mezi synonymními kodony se však tento jev snižoval (Cannarrozzi *et al.*, 2010).



**Obrázek č.5:** Translatování synonymních kodonů. Upraveno podle: Cannarrozzi *et al.* (2010)

Tímto je možná recyklace tRNA a urychlení translační rychlosti; zvýší se tedy efektivita translace. Autoři se domnívají, že právě využitá molekula tRNA se ještě příliš nevzdálila od ribosomu (neboť její pohyb je pomalejší než její opětovná aminoacylace a účast v translaci) a může se tedy rychleji opět zapojit v translačním procesu; či že tato tRNA zůstává přímo spojena s ribosomem a proto bývá opět využita (Cannarrozzi *et al.*, 2010).

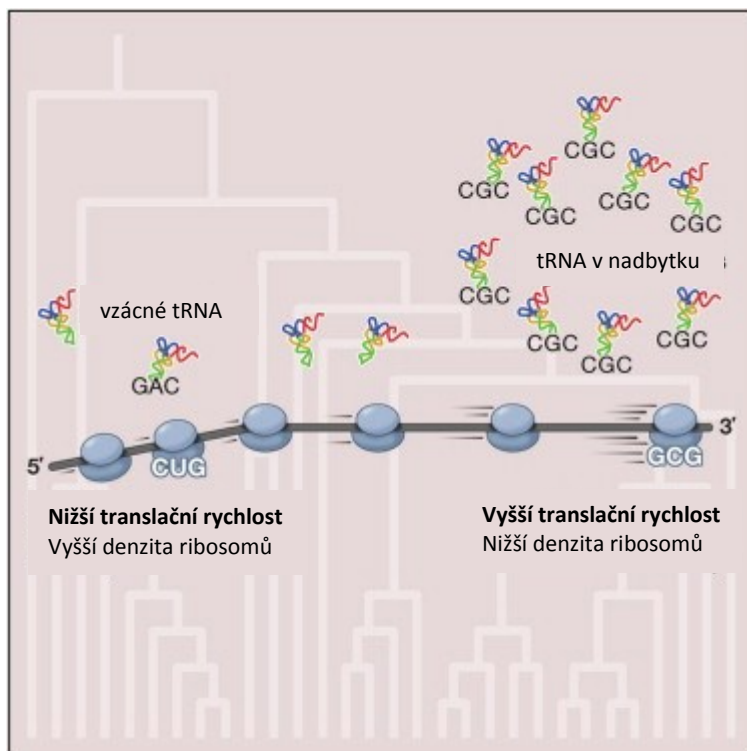
Podobný jev byl zjištěn i u *Escherichia coli* O157:H7 (kmen Sakai), což značí, že optimalizace efektivity translace probíhá i u bakterií. Podobně jako ve studii Cannarozzi *et al.* (2010), byly popsány tak zvané kodonové páry (dvojice kodonů v transkriptu, které za sebou nemusí bezprostředně následovat, ale kódují identickou aminokyselinu v proteinu). Celkově bylo studováno 773 bakteriálních genomů. V transkriptech *E. coli* byl však na rozdíl od *Saccharomyces cerevisiae* druhý kodon z páru převážně složený ze stejného tripletu bází (například UUU-UUU, UUC-UUC pro fenylalanin). Toto se potvrdilo i u ostatních zkoumaných bakteriálních kmenů. V menší míře se uplatňovaly páry kodonů, které nebyly shodné (UUU-UUC či UUC-UUU) (Shao *et al.*, 2012).

Používání kodonových párů bylo pozorováno i u archeí. Podobně jako u bakterií, také archea více využívají dvojice identických kodonů. Studie byla provedena na 122 archeálních genomech (Zhang *et al.*, 2013).

Zdá se tedy, že ve všech doménách života (prokaryota, archea i eukaryota) dochází k preferování mezi synonymními kodony, jako způsobu regulace translační efektivity.

## 6.2 Používání synonymních kodonů a efektivita translace

Tuller *et al.* (2010) zkoumali pomocí indexu tAI (viz výše) na velkém počtu osekvenovaných genomů různých organismů zastoupení jednotlivých kodonů a množství molekul tRNA, které je rozeznávají; a jak tyto faktory ovlivňují předpokládanou rychlost translace. Pro prvních 30-50 kodonů byla zjištěna účinnost translace obvykle nižší než u zbytku proteinu (u následujících kodonů došlo k 5-10% zvýšení účinnosti translace, u většiny eukaryotických organismů dále i ke konci transkriptu). Průměrná hranice, kdy došlo k popsanému jevu, byla stanovena u eukaryot na 34,5 kodonu, u prokaryot na 24 kodonů. Toto počáteční zpomalení translační rychlosti se nazývá „ramp“ a autoři ho vysvětlují jako způsob zabránění „dopravní zácpě“ ribosomů, která by snížila procesivitu (Tuller *et al.*, 2010) (Fredrick & Ibba, 2010). Přehled ukazuje **obrázek č.6.**



**Obrázek č.6:** Rychlost translace je ovlivněna použitými kodony a abundancí příslušných tRNA molekul. Upraveno podle: Tuller *et al.* (2010).

### 6.2.1 Efektivita translace - srovnání vybraných eukaryotických genomů

Kanaya *et al.* (2001) porovnávali genomy pěti eukaryotních organismů (*Schizosaccharomyces pombe*, *Caenorhabditis elegans*, *Drosophila melanogaster*, *Xenopus laevis* a *Homo sapiens*). V této studii byla zkoumána souvislost mezi frekvencí používaných kodonů a efektivitou translace, do úvahy byl vzat počet genů pro tRNA (Kanaya *et al.*, 2001). U třech z pěti zkoumaných genomů (*S. pombe*, *C. elegans*, *D. melanogaster*), zejména u více exprimovaných genů (ribosomální proteiny, histony), byla zjištěna pozitivní korelace mezi výběrem kodonů a jejich vlivem na efektivitu translace.

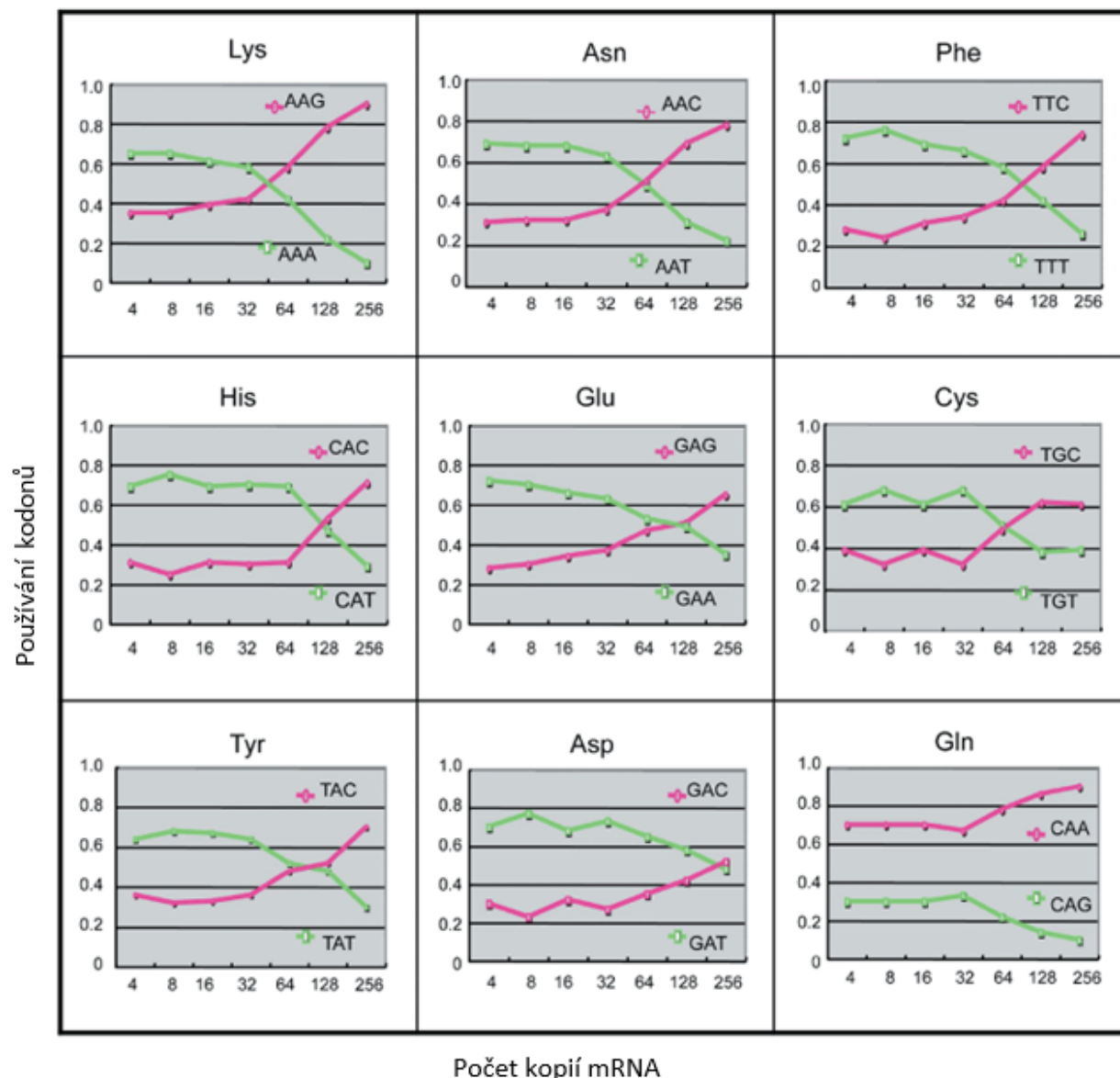
U drápatky (*X. laevis*) ani člověka se podobná závislost u těchto proteinů nepotvrdila, čehož vysvětlení by mohlo být to, že u těchto druhů je nejspíše pro různou frekvenci používání synonymních kodonů jiný důvod, než zefektivnění translace. Například u člověka by jeden z faktorů mohl být obsah GC párů v kodonech, který se liší dle genů a již na úrovni transkripce je genová exprese regulována metylací cytosinu v kódujících oblastech DNA (byly zmíněny například geny pro histony). U člověka nebyl zjištěn ani

výrazný rozdíl v počtech kopií genů pro isoakceptorové tRNA, avšak u *S. pombe* a *C. elegans* ano, v souladu s výše popsaným nálezem vyšší efektivity translace těchto výše zmíněných proteinů (Kanaya *et al.*, 2001).

### 6.3 Preference kodonů v závislosti na počtech kopií mRNA

Hiraoka *et al.* (2009) zkoumali frekvenci používání synonymních kodonů a jejich preferenci u kvasinky *Schizosaccharomyces pombe*. Při expresi ribosomálních genů byl největší rozdíl v použití kodonů oproti celogenomovému průměru pozorován u aminokyselin kódovaných dvěma kodony; nicméně rozdíly byly pozorovány i u ostatních aminokyselin. Použití kodonů s G či C na třetí pozici bylo v souladu s GC obsahem v genomu. Použití kodonů u ribosomálních genů odpovídalo kodonům, které se vyskytovaly u více exprimovaných genů (Hiraoka *et al.*, 2009).

Zjištěná frekvence používání kodonů u ribosomálních genů u „dvoukodonových“ aminokyselin *S. pombe* dle Hiraoka *et al.* (2009) je znázorněna na **obrázku č.7**. Z obrázku je jasně patrná rozdílnost v zastoupení synonymních kodonů u slabě a silně exprimovaných genů.



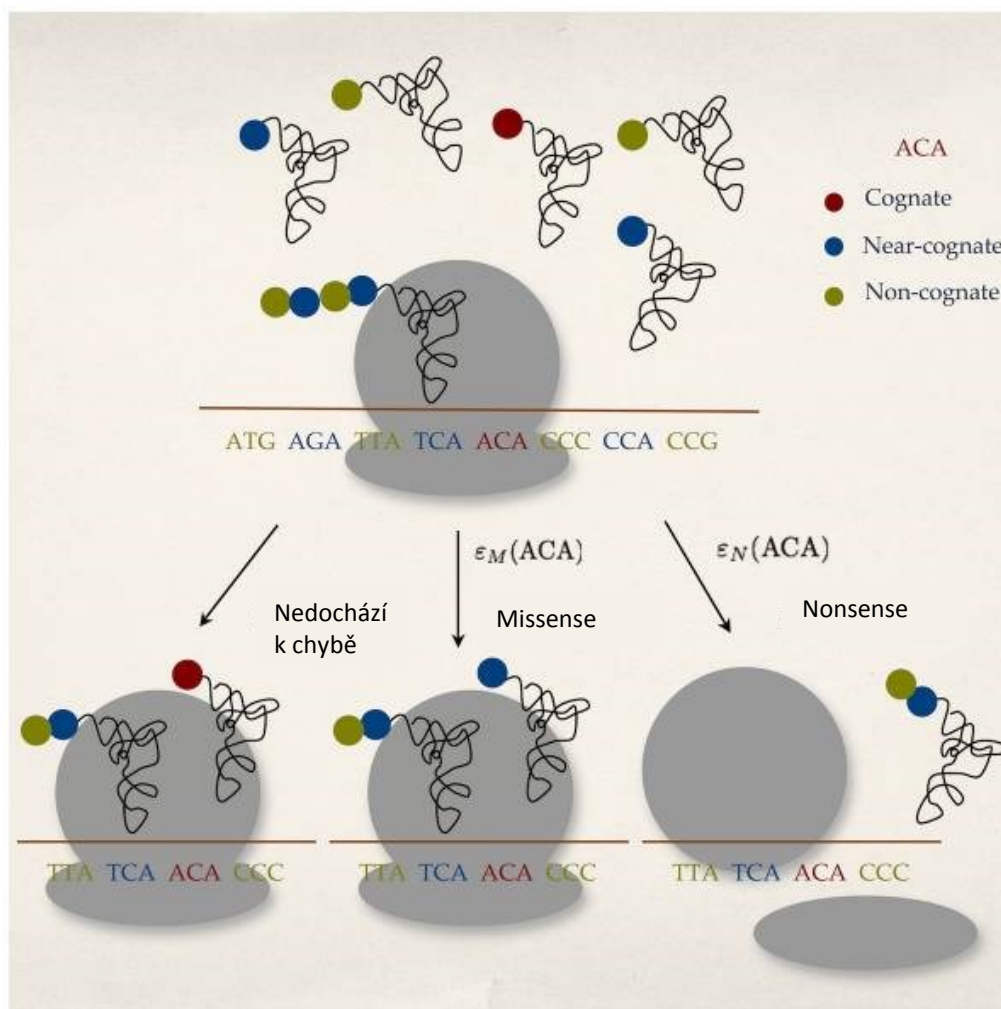
**Obrázek č.7:** Použití synonymních kodonů (frekvence použití na ose y) v závislosti na míře genové exprese (počet kopií mRNA, osa x) u kvasinky *S. pombe*. Rozdíl byl největší právě u aminokyselin kódovaných dvěma kodony, které uvádí tento přehled. Zkratky aminokyselin jsou vysvětleny výše v části Seznam použitých zkratk. Upraveno podle: Hiraoka *et al.* (2009)

#### 6.4 Vliv preference kodonů na přesnost translace

Byl zkoumán také vliv frekvence použití synonymních kodonů na přesnost translace. Stoletzki & Eyre-Walker (2007) studií genomu *Escherichia coli* zjistili, že při použití optimálnějších kodonů dochází ke

snížení četnosti chybných zařazení aminokyselin či předčasného ukončení translace (Stoletzki & Eyre-Walker, 2007).

Shah & Gilchrist (2010) provedli podobnou studii na 73 bakteriálních genomech v souvislosti s nadbytkem příslušných tRNA. Bylo zjištěno, že někdy nemusí být použití kodonů, které jsou rozeznávány větším množstvím příslušných tRNA, z hlediska přesnosti translace tou nejlepší volbou. Pokud je totiž v nadbytku i tRNA, která se liší v antikodonové sekvenci jedním nukleotidem (tzv. near-cognate tRNA), může zvýšeně docházet k chybám v zařazení aminokyseliny (missense mutace) – příklad ukazuje **obrázek č.8**.



**Obrázek č.8:** Translatování kodonu ACA. V případě čtení tRNA s odpovídajícím antikodonem (cognate tRNA) nedochází k chybě. Pokud je v nadbytku tRNA s podobným antikodonem (near-cognate tRNA; vypočteno dle tzv. „missense mutation rate“,  $\epsilon_M$ ), může dojít k missense mutaci; případně k rozpoznání

kodonu release faktory, rozpadnutí ribosomu či posunu čtecího rámce, které by vedlo k „nonsense mutaci“. Tyto procesy by měly za následek předčasné ukončení translace (s pravděpodobností „nonsense error rate“,  $\epsilon_N$ ). Příslušné výpočty uvedeny v publikaci. Upraveno podle: Shah & Gilchrist (2010)

Také byl zjištěn vztah mezi elongační rychlostí translace a s četností missense mutací. V případě nonsense mutací se však zjistilo, že jejich četnost klesala se zvýšením elongační rychlosti, což by mohlo mít na frekvenci používání synonymních kodonů významný vliv (Shah & Gilchrist, 2010).

## 7 Vliv použití synonymních kodonů na strukturu proteinu

Přesnost translace, která je daná také evoluční optimalizací frekvence použitých kodonů a jejich odpovídající tRNA, má vliv na správné složení a funkčnost proteinu. Špatně složený protein může například působit cytotoxicky (Drummond & Wilke, 2008). Objevila se tedy teorie, zda se optimální kodony více objevují na místech, která jsou z hlediska translace označována jako citlivá. Těmi jsou myšlena místa, kde by se mohla objevovat chyba při translaci a následně by se kvůli tomu protein špatně složil (T. Zhou *et al.*, 2009).

Byly zkoumány geny čtyř modelových organismů (*E. coli*, *S. cerevisiae*, *Drosophila melanogaster* a *Mus musculus*) a jejich ortology. Pomocí statistických metod (chí-kvadrát test) autoři vybrali optimální kodony u 5% nejvíce a nejméně exprimovaných genů. Tyto kodony pak analyzovali pro každý gen a každou aminokyselinu. Porovnávali jejich lokalizaci s konkrétním místem v sekvenci, jestli jsou optimální kodony použity spíše v oblastech uvnitř proteinu („buried sites“) či exponovaných místech („exposed sites“). Zjistili, že pro většinu aminokyselin byly častější optimální kodony uvnitř proteinu. U *E. coli* však u alaninu a valinu vyšel opačný výsledek. U všech čtyřech druhů nebyla zjištěna tato preference u kodonů pro cystein, kyselinu glutamovou, histidin, prolin ani tyrosin (T. Zhou *et al.*, 2009).

To, že se optimální kodony více vyskytují v místech uvnitř proteinu, se následně potvrdilo, když autoři rozšířili zkoumaný vzorek sekvencí na 25% nejvíce či nejméně exprimovaných genů, zejména u genů s vyšší expresí. Výskyt optimálnějších kodonů uvnitř proteinu by měl zabránit translačním chybám, neboť mutace v těchto místech by mohly mít výraznější dopad na strukturu proteinu po jeho složení. Mohly by tedy vést k chybnému skládání (T. Zhou *et al.*, 2009).



Vliv kodonového složení na strukturu a kotranslační skládání proteinu byl potvrzen i dalšími studiemi, například na houbě *Neurospora crassa* (Yu *et al.*, 2015) či na savčím proteinu z oční čočky, gamma-B-crystallinu (Buhr *et al.*, 2016).

### 7.1 Vliv používání vzácných kodonů na strukturu proteinu

Někteří autoři se zaměřili také na studium vzácných kodonů, tedy takových kodonů, které jsou používány s výrazně nižší frekvencí než k nim synonymní kodony. Studium mnoha prokaryotických i eukaryotických genomů bylo zjištěno, že vzácné kodony se objevují ve skupinách a nebývají umístěny jednotlivě. Toto bylo popsáno na různých genech, bez závislosti na míře jejich exprese (Clarke IV & Clark, 2008).

U *E. coli* a dalších prokaryotických organismů se ukázalo, že se vzácné kodony zvýšeně vyskytují na 3' i 5' konci kódující oblasti v porovnání se sekvencí uprostřed genu (Clarke IV & Clark, 2010). Dále bylo objeveno, že umístění těchto vzácných kodonů je do značné míry konzervované od bakterií po savce (Chaney *et al.*, 2017). Navíc záměna vzácného kodonu za optimální v těchto úsecích vede k narušení nativní hladiny exprese, struktury a funkce proteinu, jak bylo ukázáno u proteinu FRQ řídicího cirkadiální hodiny u houby *Neurospora crassa* (M. Zhou *et al.*, 2013).

Všechna tato data ukazují na fyziologickou funkci vzácných kodonů, která by mohla být vysvětlena hypotézou, podle níž dochází při čtení těchto kodonů ke krátkodobému zastavení translace, protože příslušná tRNA je v buňce málo zastoupená, což může poskytnout některým obtížně sbalitelným proteinovým doménám dostatek času na zaujmutí správné konformace (Clarke IV & Clark, 2008).

Některým kombinacím po sobě následujících kodonů se naopak organismy snaží vyhýbat. Například kombinace kodonů NNU-AAN a NNU-AGN může při posunu čtecího rámce způsobeném sklouznutím ribosomu vyústit ve STOP kodon UAA či UAG (Tats *et al.*, 2008).

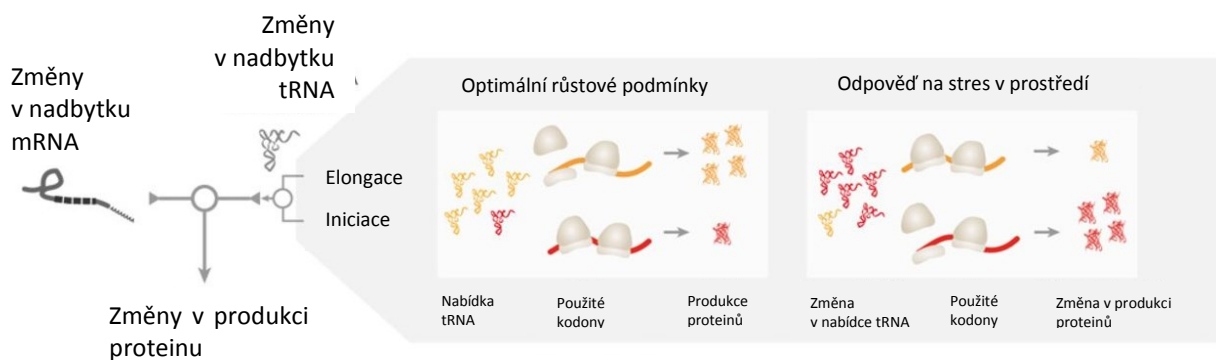
## 8 Změny genové exprese ovlivněné preferencí kodonů či nadbytkem tRNA

U kvasinky *Saccharomyces cerevisiae* bylo zjištěno, že ve stresových podmínkách, jako je například změna osmotických či oxidačních podmínek, změna teploty či diauxické přepnutí, dochází ke změně exprese tRNA (Torrent *et al.*, 2018) a modifikacím jejích nukleotidů (Deng *et al.*, 2015), čímž je ovlivněna translace a

potážmo genová exprese. Proteiny, produkované při stresu, častěji obsahují kodony, které jsou v celogenomovém měřítku vzácné.

Vlivem stresových podmínek dochází ke změně exprese jednotlivých genů pro tRNA, to znamená ke zvýšení či snížení množství různých druhů tRNA s konkrétním kodonem v buňce. Výsledné druhové složení dostupných tRNA potom více koresponduje s kodony, které jsou preferovány u stresových proteinů, což vede ke zvýšení translace těchto proteinů a adaptaci na stresové podmínky. Naopak mRNA s „optimálním“ složením kodonů (převážně geny spojené s růstem) jsou translatovány méně efektivně. Roli také hrají modifikace nukleotidů ve wobble pozici. Dochází tedy k adaptaci ke změně podmínek v dostupnosti tRNA molekul. Čím déle stresor na kvasinky působí, tím je změna v dostupnosti tRNA výraznější (Torrent *et al.*, 2018).

Přehled změn v translaci vlivem působení stresových podmínek ukazuje **obrázek č.9**.



Důsledek změn v nadbytku molekul mRNA a tRNA na proteinovou expresi

**Obrázek č.9:** Změna genové exprese jako důsledek působení stresových faktorů prostředí. Upraveno podle: Torrent *et al.* (2018).

## 9 Preference kodonů u člověka

Lavner & Kotlar (2005) ve studii genové exprese lidského genomu zjistili (pomocí matematického modelu na výpočet frekvence optimálních kodonů), že rozdílná preference kodonů se vyskytuje u více a u méně exprimovaných genů. U více exprimovaných genů jsou kodony vychýleny směrem k použití tRNA, které mají větší počet kopií genů. U méně exprimovaných genů může preference mezi synonymními kodony sloužit ke snížení translační rychlosti, jelikož jsou k translaci použity tRNA kódované menším počtem genových kopií (Lavner & Kotlar, 2005).

U člověka jako složitého mnohobuněčného organismu byla popsána odlišnost v preferenci kodonů v různých tkáních (Plotkin *et al.*, 2004) a jak již bylo zmíněno výše, i exprese isoakceptorových tRNA je ve tkáních rozdílná (Dittmar *et al.*, 2006). Sémon *et al.* (2006) však popsali, že rozdíl v používání kodonů v jednotlivých tkáních sice existuje, ale není tolik výrazný, aby selekčně ovlivnil translační proces. Možný důvod k preferování různých synonymních kodonů se zdají být různé regionální variace sekvencí (Sémon *et al.*, 2006).

Dále se například ukázalo, že preference mezi synonymními kodony je u genů na chromosomu Y celkově nízká, ale přece jenom některé kodony bývají více zastoupené (Choudhury *et al.*, 2017) (Choudhury *et al.*, 2018).

Některé vědecké skupiny se zaměřily také na výzkum vlivu nadbytku tRNA v nádorové biologii (v progresi nádorového onemocnění, například metastazování (Goodarzi *et al.*, 2016)) či vliv preference kodonů na vybraná onemocnění (například v patogenezi karcinomu ovária (Uddin *et al.*, 2019)). Tyto poznatky však přesahují náplň tématu této bakalářské práce.

## 10 Závěr

Výzkum frekvence využívání optimálních kodonů a jeho vliv na genovou expresi byl proveden na genomech řady organismů ze všech domén života. Preference kodonů (codon bias) byla zkoumána v souvislosti s množstvím dalších faktorů, jako z hlediska evolučního, celkovým složením genomu (množství GC párů), koadaptací s počtem genů pro příslušnou isoakceptorovou tRNA. Byl popsán vliv používání synonymních kodonů na efektivitu translace, růst buňky či jednobuněčných organismů. Kodonové složení má také prokázaný efekt na stabilitu molekuly mRNA a schopnost proteinu správně se složit.

Neexistuje sada kodonů, která by byla optimální univerzálně pro všechny organismy. V evoluci došlo v každém genomu k přizpůsobení frekvence používaných kodonů pro optimalizování translační efektivity, u mnohobuněčných organismů se preference mezi synonymními kodony a exprese příslušných isoakceptorových tRNA dokonce liší na úrovni jednotlivých tkání. Znalosti tohoto faktu a optimalizace kodonů pro konkrétní mikroorganismus se dá využít například při heterologní expresi proteinů mikroorganismy.

Frekvence používání určitého kodonu je úzce spjata s dostupností jeho isoakceptorové tRNA. Díky fenoménu wobble a četným modifikacím bází v rámci molekuly tRNA může být sice kodon čten více transferovými RNA, avšak v přítomnosti perfektní shody v párování kodon-antikodon je upřednostněna molekula tRNA s takovýmto antikodonem. Dostupnost isoakceptorové tRNA souvisí s počtem kopií jejich genů v genomu.

Vzácné kodony však nebyly evolučně z genomu odstraněny, neboť jejich pomalejší translaci je získán čas na správné kotranslační skládání proteinu. Vyskytují se spíše ve skupinách a méně často ojedinele. Proteiny, produkované v odpovědi na stresové podmínky prostředí, jsou kódovány také vzácnějšími kodony, avšak došlo k adaptaci na změny dostupnosti tRNA, vyvolané těmito vnějšími stresory. Také, ačkoli START kodon je stejný, mezi jednotlivými organismy se liší i frekvence používání a zastoupení STOP kodonů.

V této práci jsem shrnula dostupnou literaturu na dané téma, abych nastínila stav současného poznání. Stále ještě tento fenomén nebyl úplně objasněn. Věřím však, že se stále větším množstvím poznatků v molekulární biologii bude výzkum preference mezi synonymními kodony dále pokračovat a dojde k rozšíření poznatků a upřesnění dalších faktorů, které zatím nejsou známy.

## 11 Seznam použité literatury

- Alkatib, S., Scharff, L. B., Rogalski, M., Fleischmann, T. T., Matthes, A., Seeger, S., ... Bock, R. (2012). The Contributions of Wobbling and Superwobbling to the Reading of the Genetic Code. *PLoS Genetics*, 8(11), e1003076. <https://doi.org/10.1371/journal.pgen.1003076>
- Anderson, P., & Ivanov, P. (2014, November 28). TRNA fragments in human health and disease. *FEBS Letters*. Elsevier. <https://doi.org/10.1016/j.febslet.2014.09.001>
- Berg, O. G., & Kurland, C. G. (1997, July 25). Growth rate-optimised tRNA abundance and codon usage. *Journal of Molecular Biology*. Academic Press. <https://doi.org/10.1006/jmbi.1997.1142>
- Bernardi, G., Olofsson, B., Filipski, J., Zerial, M., Salinas, J., Cuny, G., ... Rodier, F. (1985). The mosaic genome of warm-blooded vertebrates. *Science*. <https://doi.org/10.1126/science.4001930>
- Björk, G. R., Huang, B., Persson, O. P., & Byström, A. S. (2007). A conserved modified wobble nucleoside (mcm5s2U) in lysyl-tRNA is required for viability in yeast. *RNA*, 13(8), 1245–1255. <https://doi.org/10.1261/rna.558707>
- Brule, C. E., & Grayhack, E. J. (2017, April 1). Synonymous Codons: Choose Wisely for Expression. *Trends in Genetics*. Elsevier Ltd. <https://doi.org/10.1016/j.tig.2017.02.001>
- Buhr, F., Jha, S., Thommen, M., Mittelstaet, J., Kutz, F., Schwalbe, H., ... Komar, A. A. (2016). Synonymous Codons Direct Cotranslational Folding toward Different Protein Conformations. *Molecular Cell*, 61(3), 341–351. <https://doi.org/10.1016/j.molcel.2016.01.008>
- Bulmer, M. (1991). The selection-mutation-drift theory of synonymous codon usage. *Genetics*, 129(3), 897–907.
- Bulmer, Michael. (1987). Coevolution of codon usage and transfer RNA abundance. *Nature*, 325(6106), 728–730. <https://doi.org/10.1038/325728a0>
- Cannarozzi, G., Schraudolph, N. N., Faty, M., von Rohr, P., Friberg, M. T., Roth, A. C., ... Barral, Y. (2010). A role for codon order in translation dynamics. *Cell*, 141(2), 355–367. <https://doi.org/10.1016/j.cell.2010.02.036>
- Cantara, W. A., Crain, P. F., Rozenski, J., McCloskey, J. A., Harris, K. A., Zhang, X., ... Agris, P. F. (2011). The RNA Modification Database, RNAMDB: 2011 update. *Nucleic Acids Research*, 39(Database issue), D195-201. <https://doi.org/10.1093/nar/gkq1028>

- Chaney, J. L., Steele, A., Carmichael, R., Rodriguez, A., Specht, A. T., Ngo, K., ... Clark, P. L. (2017). Widespread position-specific conservation of synonymous rare codons within coding sequences. *PLoS Computational Biology*, *13*(5). <https://doi.org/10.1371/journal.pcbi.1005531>
- Chen, S. L., Lee, W., Hottes, A. K., Shapiro, L., & McAdams, H. H. (2004). Codon usage between genomes is constrained genome-wide mutational processes. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(10), 3480–3485. <https://doi.org/10.1073/pnas.0307827100>
- Chin, J. X., Chung, B. K.-S., & Lee, D.-Y. (2014). Codon Optimization OnLine (COOL): a web-based multi-objective optimization platform for synthetic gene design. *Bioinformatics (Oxford, England)*, *30*(15), 2210–2212. <https://doi.org/10.1093/bioinformatics/btu192>
- Choudhury, M. N., Uddin, A., & Chakraborty, S. (2018). Nucleotide composition and codon usage bias of SRY gene. *Andrologia*, *50*(1). <https://doi.org/10.1111/and.12787>
- Chu, D., Kazana, E., Bellanger, N., Singh, T., Tuite, M. F., & von der Haar, T. (2014). Translation elongation can control translation initiation on eukaryotic mRNAs. *The EMBO Journal*, *33*(1), 21–34. <https://doi.org/10.1002/emj.201385651>
- Clark, B. F. C. (2006). The crystal structure of tRNA. *Journal of Biosciences*, *31*(4), 453–457. <https://doi.org/10.1007/bf02705184>
- Clarke IV, T. F., & Clark, P. L. (2008). Rare codons cluster. *PLoS ONE*, *3*(10). <https://doi.org/10.1371/journal.pone.0003412>
- Clarke IV, T. F., & Clark, P. L. (2010). Increased incidence of rare codon clusters at 5' and 3' gene termini: Implications for function. *BMC Genomics*, *11*(1), 118. <https://doi.org/10.1186/1471-2164-11-118>
- Crick, F. H. C. (1966). Codon—anticodon pairing: The wobble hypothesis. *Journal of Molecular Biology*, *19*(2), 548–555. [https://doi.org/10.1016/S0022-2836\(66\)80022-0](https://doi.org/10.1016/S0022-2836(66)80022-0)
- Daniel, E., Onwukwe, G. U., Wierenga, R. K., Quaggin, S. E., Vainio, S. J., & Krause, M. (2015). ATGme: Open-source web application for rare codon identification and custom DNA sequence optimization. *BMC Bioinformatics*, *16*(1), 303. <https://doi.org/10.1186/s12859-015-0743-5>
- Deng, W., Babu, I. R., Su, D., Yin, S., Begley, T. J., & Dedon, P. C. (2015). Trm9-Catalyzed tRNA Modifications Regulate Global Protein Expression by Codon-Biased Translation. *PLoS Genetics*, *11*(12). <https://doi.org/10.1371/journal.pgen.1005706>

- Dittmar, K. A., Goodenbour, J. M., & Pan, T. (2006). Tissue-specific differences in human tRNA expression. *PLoS Genetics*, 2(12), 2107–2115. <https://doi.org/10.1371/journal.pgen.0020221>
- dos Reis, M., Savva, R., & Wernisch, L. (2004). Solving the riddle of codon usage preferences: A test for translational selection. *Nucleic Acids Research*, 32(17), 5036–5044. <https://doi.org/10.1093/nar/gkh834>
- Drummond, D. A., & Wilke, C. O. (2008). Mistranslation-Induced Protein Misfolding as a Dominant Constraint on Coding-Sequence Evolution. *Cell*, 134(2), 341–352. <https://doi.org/10.1016/j.cell.2008.05.042>
- Elena, C., Ravasi, P., Castelli, M. E., Peirú, S., & Menzella, H. G. (2014, February 4). Expression of codon optimized genes in microbial systems: Current industrial applications and perspectives. *Frontiers in Microbiology*. Frontiers Research Foundation. <https://doi.org/10.3389/fmicb.2014.00021>
- Fredrick, K., & Ibba, M. (2010, April 16). How the sequence of a gene can tune its translation. *Cell*. <https://doi.org/10.1016/j.cell.2010.03.033>
- Galtier, N., Piganeau, G., Mouchiroud, D., & Duret, L. (2001). GC-content evolution in mammalian genomes: The biased gene conversion hypothesis. *Genetics*.
- Goodarzi, H., Nguyen, H. C. B., Zhang, S., Dill, B. D., Molina, H., & Tavazoie, S. F. (2016). Modulated expression of specific tRNAs drives gene expression and cancer progression. *Cell*, 165(6), 1416–1427. <https://doi.org/10.1016/j.cell.2016.05.046>
- Gustafsson, C., Govindarajan, S., & Minshull, J. (2004, July 1). Codon bias and heterologous protein expression. *Trends in Biotechnology*. Elsevier Current Trends. <https://doi.org/10.1016/j.tibtech.2004.04.006>
- Han, L., & Phizicky, E. M. (2018, October 1). A rationale for tRNA modification circuits in the anticodon loop. *RNA*. Cold Spring Harbor Laboratory Press. <https://doi.org/10.1261/rna.067736.118>
- Hartman, H., & Smith, T. F. (2014). The evolution of the ribosome and the genetic code. *Life*, 4(2), 227–249. <https://doi.org/10.3390/life4020227>
- Hershberg, R., & Petrov, D. A. (2008). Selection on Codon Bias. *Annual Review of Genetics*, 42(1), 287–299. <https://doi.org/10.1146/annurev.genet.42.110807.091442>
- Hershberg, R., & Petrov, D. A. (2009). General rules for optimal codon choice. *PLoS Genetics*, 5(7),

e1000556. <https://doi.org/10.1371/journal.pgen.1000556>

- Hershberg, R., & Petrov, D. A. (2012). On the Limitations of Using Ribosomal Genes as References for the Study of Codon Usage: A Rebuttal. *PLoS ONE*, 7(12), e49060. <https://doi.org/10.1371/journal.pone.0049060>
- Higgs, P. G., & Ran, W. (2008). Coevolution of codon usage and tRNA genes leads to alternative stable states of biased codon usage. *Molecular Biology and Evolution*, 25(11), 2279–2291. <https://doi.org/10.1093/molbev/msn173>
- Hiraoka, Y., Kawamata, K., Haraguchi, T., & Chikashige, Y. (2009). Codon usage bias is correlated with gene expression levels in the fission yeast *Schizosaccharomyces pombe*. *Genes to Cells*, 14(4), 499–509. <https://doi.org/10.1111/j.1365-2443.2009.01284.x>
- Ikemura, T. (1981). Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes: A proposal for a synonymous codon choice that is optimal for the *E. coli* translational system. *Journal of Molecular Biology*, 151(3), 389–409. [https://doi.org/10.1016/0022-2836\(81\)90003-6](https://doi.org/10.1016/0022-2836(81)90003-6)
- Ingolia, N. T., Ghaemmaghami, S., Newman, J. R. S., & Weissman, J. S. (2009). Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science*, 324(5924), 218–223. <https://doi.org/10.1126/science.1168978>
- Johansson, M. J. O., Esberg, A., Huang, B., Bjork, G. R., & Bystrom, A. S. (2008). Eukaryotic Wobble Uridine Modifications Promote a Functionally Redundant Decoding System. *Molecular and Cellular Biology*, 28(10), 3301–3312. <https://doi.org/10.1128/mcb.01542-07>
- Kanaya, S., Yamada, Y., Kinouchi, M., Kudo, Y., & Ikemura, T. (2001). Codon usage and tRNA genes in eukaryotes: Correlation of codon usage diversity with translation efficiency and with CG-dinucleotide usage as assessed by multivariate analysis. *Journal of Molecular Evolution*, 53(4–5), 290–298. <https://doi.org/10.1007/s002390010219>
- Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution*, 16(2), 111–120. <https://doi.org/10.1007/BF01731581>
- Knight, R. D., Freeland, S. J., & Landweber, L. F. (2001). A simple model based on mutation and selection



- explains trends in codon and amino-acid usage and GC composition within and across genomes. *Genome Biology*, 2(4). <https://doi.org/10.1186/gb-2001-2-4-research0010>
- Lajoie, M. J., Rovner, A. J., Goodman, D. B., Aerni, H. R., Haimovich, A. D., Kuznetsov, G., ... Isaacs, F. J. (2013). Genomically recoded organisms expand biological functions. *Science*, 342(6156), 357–360. <https://doi.org/10.1126/science.1241459>
- Lavner, Y., & Kotlar, D. (2005). Codon bias as a factor in regulating expression via translation rate in the human genome. *Gene*, 345(1 SPEC. ISS.), 127–138. <https://doi.org/10.1016/j.gene.2004.11.035>
- Lorenz, C., Lünse, C. E., & Mörl, M. (2017, June 1). Trna modifications: Impact on structure and thermal adaptation. *Biomolecules*. MDPI AG. <https://doi.org/10.3390/biom7020035>
- Mauro, V. P., & Chappell, S. A. (2014, November 1). A critical analysis of codon optimization in human therapeutics. *Trends in Molecular Medicine*. Elsevier Ltd. <https://doi.org/10.1016/j.molmed.2014.09.003>
- Mignon, C., Mariano, N., Stadthagen, G., Lugari, A., Lagoutte, P., Donnat, S., ... Werle, B. (2018, May 1). Codon harmonization – going beyond the speed limit for protein expression. *FEBS Letters*. Wiley Blackwell. <https://doi.org/10.1002/1873-3468.13046>
- Nakamura, Y. (2000). Codon usage tabulated from international DNA sequence databases: status for the year 2000. *Nucleic Acids Research*, 28(1), 292–292. <https://doi.org/10.1093/nar/28.1.292>
- Nath Choudhury, M., Uddin, A., & Chakraborty, S. (2017). Codon usage bias and its influencing factors for Y-linked genes in human. *Computational Biology and Chemistry*, 69, 77–86. <https://doi.org/10.1016/j.compbiolchem.2017.05.005>
- Nobles, K. N., Yarian, C. S., Liu, G., Guenther, R. H., & Agris, P. F. (2002). Highly conserved modified nucleosides influence Mg<sup>2+</sup>-dependent tRNA folding. *Nucleic Acids Research*, 30(21), 4751–4760. <https://doi.org/10.1093/nar/gkf595>
- Novoa, E. M., & Ribas de Pouplana, L. (2012, November). Speeding with control: Codon usage, tRNAs, and ribosomes. *Trends in Genetics*. <https://doi.org/10.1016/j.tig.2012.07.006>
- Pak, D., Root-Bernstein, R., & Burton, Z. F. (2017). tRNA structure and evolution and standardization to the three nucleotide genetic code. *Transcription*, 8(4), 205–219. <https://doi.org/10.1080/21541264.2017.1318811>

- Pan, T. (2018, April 1). Modifications and functional genomics of human transfer RNA. *Cell Research*. Nature Publishing Group. <https://doi.org/10.1038/s41422-018-0013-y>
- Panicker, I. S., Browning, G. F., & Markham, P. F. (2015). The effect of an alternate start codon on heterologous expression of a PhoA fusion protein in mycoplasma gallisepticum. *PLoS ONE*, *10*(5), e0127911. <https://doi.org/10.1371/journal.pone.0127911>
- Paul, P., Malakar, A. K., & Chakraborty, S. (2018). Codon usage vis-a-vis start and stop codon context analysis of three dicot species. *Journal of Genetics*, *97*(1), 97–107. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/29666329>
- Pechmann, S., & Frydman, J. (2013). Evolutionary conservation of codon optimality reveals hidden signatures of cotranslational folding. *Nature Structural and Molecular Biology*, *20*(2), 237–243. <https://doi.org/10.1038/nsmb.2466>
- Percudani, R., Pavesi, A., & Ottonello, S. (1997). Transfer RNA gene redundancy and translational selection in *Saccharomyces cerevisiae*. *Journal of Molecular Biology*, *268*(2), 322–330. <https://doi.org/10.1006/jmbi.1997.0942>
- Plotkin, J. B., Robins, H., & Levine, A. J. (2004). Tissue-specific codon usage and the expression of human genes. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(34), 12588–12591. <https://doi.org/10.1073/pnas.0404957101>
- Presnyak, V., Alhusaini, N., Chen, Y. H., Martin, S., Morris, N., Kline, N., ... Collier, J. (2015). Codon optimality is a major determinant of mRNA stability. *Cell*, *160*(6), 1111–1124. <https://doi.org/10.1016/j.cell.2015.02.029>
- Puigbò, P., Guzmán, E., Romeu, A., & Garcia-Vallvé, S. (2007). OPTIMIZER: a web server for optimizing the codon usage of DNA sequences. *Nucleic Acids Research*, *35*(Web Server issue), W126-31. <https://doi.org/10.1093/nar/gkm219>
- Quax, T. E. F., Claassens, N. J., Söll, D., & van der Oost, J. (2015, July 16). Codon Bias as a Means to Fine-Tune Gene Expression. *Molecular Cell*. Cell Press. <https://doi.org/10.1016/j.molcel.2015.05.035>
- Sabi, R., & Tuller, T. (2014). Modelling the efficiency of codon-tRNA interactions based on codon usage bias. *DNA Research: An International Journal for Rapid Publication of Reports on Genes and Genomes*, *21*(5), 511–526. <https://doi.org/10.1093/dnares/dsu017>

- Sánchez, R., Grau, R., & Morgado, E. (2006). A novel Lie algebra of the genetic code over the Galois field of four DNA bases. *Mathematical Biosciences*, 202(1), 156–174. <https://doi.org/10.1016/j.mbs.2006.03.017>
- Sémon, M., Lobry, J. R., & Duret, L. (2006). No evidence for tissue-specific adaptation of synonymous codon usage in humans. *Molecular Biology and Evolution*, 23(3), 523–529. <https://doi.org/10.1093/molbev/msj053>
- Shah, P., & Gilchrist, M. A. (2010). Effect of Correlated tRNA abundances on translation errors and evolution of codon usage bias. *PLoS Genetics*, 6(9), e1001128. <https://doi.org/10.1371/journal.pgen.1001128>
- Shao, Z. Q., Zhang, Y. M., Feng, X. Y., Wang, B., & Chen, J. Q. (2012). Synonymous codon ordering: A subtle but prevalent strategy of bacteria to improve translational efficiency. *PLoS ONE*, 7(3), e33547. <https://doi.org/10.1371/journal.pone.0033547>
- Sharp, P. M., & Li, W. H. (1987). The codon adaptation index—a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Research*, 15(3), 1281–1295. <https://doi.org/10.1093/nar/15.3.1281>
- Sharp, P. M., Tuohy, T. M. F., & Mosurski, K. R. (1986). Codon usage in yeast: Cluster analysis clearly differentiates highly and lowly expressed genes. *Nucleic Acids Research*, 14(13), 5125–5143. <https://doi.org/10.1093/nar/14.13.5125>
- Sprinzi, M., Horn, C., Brown, M., Loudovltch, A., & Steinberg, S. (1998). Compilation of tRNA sequences and sequences of tRNA genes. *Nucleic Acids Research*, 26(1), 148–153. <https://doi.org/10.1093/nar/26.1.148>
- Stoletzki, N., & Eyre-Walker, A. (2007). Synonymous codon usage in Escherichia coli: selection for translational accuracy. *Molecular Biology and Evolution*, 24(2), 374–381. <https://doi.org/10.1093/molbev/msl166>
- Sun, J., Chen, M., Xu, J., & Luo, J. (2005). Relationships among stop codon usage bias, its context, isochores, and gene expression level in various eukaryotes. *Journal of Molecular Evolution*, 61(4), 437–444. <https://doi.org/10.1007/s00239-004-0277-3>
- Tats, A., Tenson, T., & Remm, M. (2008). Preferred and avoided codon pairs in three domains of life. *BMC*

*Genomics*, 9, 463. <https://doi.org/10.1186/1471-2164-9-463>

- Torrent, M., Chalancon, G., De Groot, N. S., Wuster, A., & Madan Babu, M. (2018). Cells alter their tRNA abundance to selectively regulate protein synthesis during stress conditions. *Science Signaling*, 11(546). <https://doi.org/10.1126/scisignal.aat6409>
- Trotta, E. (2013). Selection on codon bias in yeast: a transcriptional hypothesis. *Nucleic Acids Research*, 41(20), 9382–9395. <https://doi.org/10.1093/nar/gkt740>
- Trotta, E. (2016). Selective forces and mutational biases drive stop codon usage in the human genome: A comparison with sense codon usage. *BMC Genomics*, 17(1). <https://doi.org/10.1186/s12864-016-2692-4>
- Tuller, T., Carmi, A., Vestsigian, K., Navon, S., Dorfan, Y., Zaborske, J., ... Pilpel, Y. (2010). An evolutionarily conserved mechanism for controlling the efficiency of protein translation. *Cell*, 141(2), 344–354. <https://doi.org/10.1016/j.cell.2010.03.031>
- Uddin, A., Paul, N., & Chakraborty, S. (2019). The codon usage pattern of genes involved in ovarian cancer. *Annals of the New York Academy of Sciences*, 1440(1), 67–78. <https://doi.org/10.1111/nyas.14019>
- Urbonavicius, J., Qian, Q., Durand, J. M., Hagervall, T. G., & Björk, G. R. (2001). Improvement of reading frame maintenance is a common function for several tRNA modifications. *The EMBO Journal*, 20(17), 4863–4873. <https://doi.org/10.1093/emboj/20.17.4863>
- Vicario, S., Moriyama, E. N., & Powell, J. R. (2007). Codon usage in twelve species of *Drosophila*. *BMC Evolutionary Biology*, 7(1), 226. <https://doi.org/10.1186/1471-2148-7-226>
- Wald, N., Alroy, M., Botzman, M., & Margalit, H. (2012). Codon usage bias in prokaryotic pyrimidine-ending codons is associated with the degeneracy of the encoded amino acids. *Nucleic Acids Research*, 40(15), 7074–7083. <https://doi.org/10.1093/nar/gks348>
- Wang, B., Shao, Z. Q., Xu, Y., Liu, J., Liu, Y., Hang, Y. Y., & Chen, J. Q. (2011). Optimal codon identities in bacteria: Implications from the conflicting results of two different methods. *PLoS ONE*, 6(7), e22714. <https://doi.org/10.1371/journal.pone.0022714>
- Weixlbaumer, A., Murphy IV, F. V., Dziergowska, A., Malkiewicz, A., Vendeix, F. A. P., Agris, P. F., & Ramakrishnan, V. (2007). Mechanism for expanding the decoding capacity of transfer RNAs by modification of uridines. *Nature Structural and Molecular Biology*, 14(6), 498–502.

<https://doi.org/10.1038/nsmb1242>

- Yamao, F., Muto, A., Kawauchi, Y., Iwami, M., Iwagami, S., Azumi, Y., & Osawa, S. (1985). UGA is read as tryptophan in *Mycoplasma capricolum*. *Proceedings of the National Academy of Sciences of the United States of America*, *82*(8), 2306–2309. <https://doi.org/10.1073/pnas.82.8.2306>
- Yu, C. H., Dang, Y., Zhou, Z., Wu, C., Zhao, F., Sachs, M. S., & Liu, Y. (2015). Codon Usage Influences the Local Rate of Translation Elongation to Regulate Co-translational Protein Folding. *Molecular Cell*, *59*(5), 744–754. <https://doi.org/10.1016/j.molcel.2015.07.018>
- Zhang, Y. M., Shao, Z. Q., Yang, L. T., Sun, X. Q., Mao, Y. F., Chen, J. Q., & Wang, B. (2013). Non-random arrangement of synonymous codons in archaea coding sequences. *Genomics*, *101*(6), 362–367. <https://doi.org/10.1016/j.ygeno.2013.04.008>
- Zhou, M., Guo, J., Cha, J., Chae, M., Chen, S., Barral, J. M., ... Liu, Y. (2013). Non-optimal codon usage affects expression, structure and function of clock protein FRQ. *Nature*, *494*(7439), 111–115. <https://doi.org/10.1038/nature11833>
- Zhou, T., Weems, M., & Wilke, C. O. (2009). Translationally optimal codons associate with structurally sensitive sites in proteins. *Molecular Biology and Evolution*, *26*(7), 1571–1580. <https://doi.org/10.1093/molbev/msp070>